


```

:      ORGANISM: Nicotiana tabacum/Cladosporium fulvum
:
:      STRAIN: Cladosporium fulvum race 2,5
:
:      IMMEDIATE SOURCE:
:
:      CLONE: SP:AVR4
:
US-08-945-983-3

```

Query Match	31.28;	Score 61.4;	DB 4;	Length 484;
Best Local Similarity	67.78;	Pred. No. 2e-10;		
Matches 86; Conservative	0;	Mismatches 41;	Indels 0;	Gaps 0;

QY	130	actgaa	136
Db	123	CATGCA	129

RESULT 2
US-08-945-983-5/c

```

: GENERAL INFORMATION:
: APPLICANT: Thomas, Colwyn M
: APPLICANT: Balint-Kurtli, Peter J
: APPLICANT: Jones, David A
: APPLICANT: Jones, Jonathan DG
: TITLE OF INVENTION: Plant pathogen resistance genes and uses
: TITLE OF INVENTION: thereof
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:

```

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/945,983
8  FILING DATE: 12-NOV-1997

```

1 PRIOR APPLICATION DATA:
2 APPLICATION NUMBER: PCT/GB96/01155
3 FILING DATE: 13-MAY-1996
4 PRIOR APPLICATION DATA:
5 APPLICATION NUMBER: GB 9509575.8
6 FILING DATE: 11-MAY-1995
7 ATTORNEY/AGENT INFORMATION:

```

; INFORMATION FOR SEQ ID NO:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 484 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: double
;
; TOPOLOGY: linear
;
US-08-945-983-5

```

Query Match	31.2%;	Score 61.4;	DB 4;	Length 48;
Best Local Similarity	67.7%;	Pred. No. 2e-10;		
Matches 86; Conservative	0;	Mismatches 41;	Indels 0;	Gaps 0;

Qy	10	caatggatcttcgtatcttcctcagcttccaaactcttcctcgtctcctctcttc	69
Db	482	CGATGGGATTTTCTCTTTTCAACAATTCGCTCAATTTCTGTGCTCAACATCTCT	423
Qy	70	tctctctctgtactcactctctgctgacggtctccagaagccagctgtccaatcact	129
Db	422	TATTCCTAATATATATCCACACTCTGGCGGTCCAAAGCCCAAAATCTCAACATTCACAA	365
Qy	130	actgcac	136
Db	362	CATGCAC	356

RESULT 3
US-08-181-271A-9
; Sequence 9, Application US/08181271A
; Patent No. 5614395

1 GENERAL INFORMATION:
2 APPLICANT: Ryals, John A.
3 APPLICANT: Alexander, Danny C.
4 APPLICANT: Beck, James J.
5 APPLICANT: Duesing, John H.
6 APPLICANT: Friedrich, Leslie B.
7 APPLICANT: Goodman, Robert M.
8 APPLICANT: Harms, Christian
9 APPLICANT: Melns, Jr., Frederick
10 APPLICANT: Montoya, Alice
11 APPLICANT: Moyer, Mary B.
12 APPLICANT: Neuhaus, Jean-Marc
13 APPLICANT: Payne, George B.
14 APPLICANT: Sperison, Christoph
15 APPLICANT: Stinson, Jeffrey R.
16 APPLICANT: Uknes, Scott J.

```

1      TITLE OF INVENTION:  CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
2
3      TITLE OF INVENTION:
4
5      NUMBER OF SEQUENCES:  DNA SEQUENCES AND USES THEREOF
6
7      CORRESPONDENCE ADDRESS:  106
8
9

```

```

:
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patulinin Release #1.0, Version #1.25
:

```

1 PRIOR APPLICATION DATA:
 2 APPLICATION NUMBER: US 08/093,301
 3 FILING DATE: 16-JUL-1993
 4 PRIOR APPLICATION DATA:
 5 APPLICATION NUMBER: US 07/937,197
 6 FILING DATE: 6-NOV-1992
 7 PRIOR APPLICATION DATA:
 8 APPLICATION NUMBER: US 07/678,378
 9 FILING DATE: 1-APR-1991
 10 PRIOR APPLICATION DATA:
 11 APPLICATION NUMBER: US 07/305,566
 12 FILING DATE: 6-FEB-1989
 13 PRIOR APPLICATION DATA:
 14 APPLICATION NUMBER: US 07/165,657


```

GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Helms, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
City: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,043
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:

```

```

1 APPLICATION NUMBER: US 07/368,672
2 FILING DATE: 20-JUN-1989
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US 07/329,018
5 FILING DATE: 24-MAR-1989
6 PRIOR APPLICATION DATA:
7 APPLICATION NUMBER: US 08/045,957
8 FILING DATE: 12-APR-1993
9 ATTORNEY/AGENT INFORMATION:
10 NAME: Elmer, James Scott
11 REGISTRATION NUMBER: 36,129
12 REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: (919)541-8614
15 TELEFAX: (919)541-8689
16 INFORMATION FOR SEQ ID NO: 9:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 809 base pairs
19 TYPE: nucleic acid
20 STRANDEDNESS: single
21 TOPOLOGY: linear
22 MOLECULE TYPE: cDNA
23 US-08-449-043-9
24
25 Query Match 29.3% Score 57.8: DB 1: Length 809:
26 Best Local Similarity 76.3%: Pred. No. 3.3e-09;
27 Matches 71: Conservative 0; Mismatches 22; Indels 0; Gaps 0.
28
29 Oy 9 gccatgggttttggtcttctctcagcttcacatcttctctgtgtctactctt 68
30 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
31 Db 27 gcatgggatttgctcttcttccacaaatggcccttcattcttctgtcctacacttctc 86
32 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
33 Oy 69 cttctctgtatctcaccatcttgcctgcccgtgc 101
34 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
35 Db 87 ttattcctagatatatccacactcttgcctgtcc 119
36
37 RESULT 7
38 US-08-456-265A-9
39 ; Sequence 9, Application US/08456265A
40 ; Patent No. 5767369
41 ; GENERAL INFORMATION:
42 ; APPLICANT: Alexander, Danny C.
43 ; APPLICANT: Ryals, John A.
44 ; APPLICANT: Goodman, Robert M.
45 ; APPLICANT: Stinson, Jeffrey R.
46 ; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
47 ; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
48 ; NUMBER OF SEQUENCES: 111
49 ; CORRESPONDENCE ADDRESSES:
50 ; ADDRESSEE: CIBA-GEIGY Corporation
51 ; STREET: 520 White Plains Road, P.O. Box 2005
52 ; CITY: Tarrytown
53 ; STATE: New York
54 ; COUNTRY: USA
55 ; ZIP: 10591
56
57 COMPUTER READABLE FORM:
58 MEDIUM TYPE: Floppy disk
59 COMPUTER: IBM PC compatible
60 OPERATING SYSTEM: PC-DOS/MS-DOS
61 SOFTWARE: patentin Release #1.0, Version #1.25
62 CURRENT APPLICATION DATA:
63 APPLICATION NUMBER: US/08/456,265A
64 FILING DATE: 31-MAY-95
65 CLASSIFICATION: 435
66 PRIOR APPLICATION DATA:
67 APPLICATION NUMBER: US 08/181,271
68 FILING DATE: 13-JAN-1994
69 PRIOR APPLICATION DATA:
70 APPLICATION NUMBER: US 08/093,301
71 FILING DATE: 16-JUL-1993
72 PRIOR APPLICATION DATA:

```


FILING DATE: 20-JUN-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 809 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-455-416-9

Query Match 29.3%; Score 57.8; DB 1; Length 809;
Best Local Similarity 76.3%; Pred. No. 3.3e-09;
Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 9 gccatgggttcgtcttcctcagcttcacatcttcctctgtgtactctct 68
DB 27 gtcacggatgttctcttccacacatgcttcacatgttcttgcctacactctc 86
QY 69 ctttccttgatcttcacacgtgcgtgc 101
DB 87 ttattcctagtatattccacactcttcgctgc 119

RESULT 9
US-08-455-244-9
Sequence 9, Application US/08455244
Patent No. 5789214
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Spertson, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,244
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 809 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-455-244-9

Query Match 29.3%; Score 57.8; DB 1; Length 809;
Best Local Similarity 76.3%; Pred. No. 3.3e-09;
Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 9 gccatgggttcgtcttcctcagcttcacatcttcctctgtgtactctct 68
DB 27 gtcacggatgttctcttccacacatgcttcacatgttcttgcctacactctc 86

CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/457,364
 FILING DATE: 31-MAY-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JAN-94
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/425,504
 FILING DATE: 20-OCT-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,506
 FILING DATE: 6-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,122
 FILING DATE: 27-SEP-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/580,431
 FILING DATE: 7-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8614
 TELEFAX: (919)541-8689
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 809 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: CDNA
 US-08-457-364-9
 Query Match 29.3%; Score 57.8; DB 2; Length 809;
 Best Local Similarity 76.3%; Pred. No. 3.3e-09;
 Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
 QY 9 gccatgggttcgttccttcctcagctcattcttccttcgtgtctctctt 68
 Db 27 gcatgggttcgttccttcctcagctcattcttccttcgtgtctctctt 86
 QY 69 cttctcgtgactcctcactccttcgctgac 101
 Db 87 ttatctcgtgactcctcactccttcgctgac 119
 RESULT 12
 US-08-456-262-9
 Sequence 9, Application US/08456262
 Patent No. 5851766
 GENERAL INFORMATION:
 APPLICANT: Ryals, John A.
 APPLICANT: Alexander, Danny C.
 APPLICANT: Beck, James J.
 APPLICANT: Duesing, John H.
 APPLICANT: Friedrich, Leslie B.
 APPLICANT: Goodman, Robert M.
 APPLICANT: Harms, Christian
 APPLICANT: Melns, Jr., Frederick
 APPLICANT: Montoya, Alice
 APPLICANT: Moyer, Mary B.
 APPLICANT: Neuhaus, Jean-Marc
 APPLICANT: Payne, George B.
 APPLICANT: Sperison, Christoph
 APPLICANT: Stinson, Jeffrey R.
 APPLICANT: Unnes, Scott J.
 APPLICANT: Ward, Eric R.
 TITLE OF INVENTION: Williams, Shericea C.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/456,262
 FILING DATE: 31-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JAN-94
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:

Patent No. 5942662
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Harms, Christian
APPLICANT: Friedrich, Leslie
APPLICANT: Beck, James
APPLICANT: Uknes, Scott
APPLICANT: Ward, Eric
TITLE OF INVENTION: INDUCIBLE HERBICIDE RESISTANCE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESS: No. 594262artls Corporation
STREET: 3054 Cornwells Road, P.O. Box 12257
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,217
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,364
FILING DATE: 31-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/181,271
FILING DATE: 13-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957

FILED DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38, 241
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727/DIV5/CONF
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689 9:
INFORMATION FOR SEQ ID NO:
SEQUENCE CHARACTERISTICS:
LENGTH: 809 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-971-217-9
Query Match 29.3%; Score 57.8; DB 2; Length 809;
Best Local Similarity 76.3%; Pred. No. 3.3e-05;
Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 9 gccatgggttgcgtcttctcctcagcttcacatcttctcttcttgcctacattctt 68
Db 27 gtcattgggatttctctcttcttccacattgaccttcttcttcttcttcttctc 86
QY 69 ctttccttgatgactctcactcactcttgcctgccc 101
Db 87 ttattcttattatattatccacattgaccttgcctgcc 119

Search completed: May 12, 2002, 20:14:37
Job time: 7194 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 20:20:39 ; Search time 333.08 Seconds
(without alignments)

1015,468 Million cell updates/sec

Title: US-09-554-024-5

Perfect score: 197
Sequence: 1 ggcgtgcagccatgggttc.....gcgacgtgtcgcagcgtac 197

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_012802:*

1:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	197	100.0	197	20	AAK59774
2	117.4	59.6	211	20	AAK29729
3	103	52.3	106	20	AAK59773
4	103	52.3	106	20	AAK29728
5	101.6	51.6	120	21	AAZ45694
6	96	48.7	98	20	AAK59772
7	93.4	47.4	261	21	AAZ45693
8	73.4	37.3	75	20	AAK29730
9	67.4	34.2	72	20	AAK29731

10	63	32.0	63	20	AAK59771
11	61.4	31.2	484	17	AAK06310
12	61.4	31.2	484	17	AAK47876
13	58.4	29.6	107	22	AAK10075
14	57.8	29.3	809	11	AAK06180
15	57.8	29.3	809	11	AAK62805
16	57.8	29.3	809	20	AAK72997
17	57.8	29.3	1260	11	AAK04693
18	57.8	29.3	1363	10	AAK91026
19	57.8	29.3	2038	10	AAK90367
20	57.8	29.3	2038	11	AAK03663
21	57.8	29.3	2038	20	AAK62787
22	57.8	29.3	2038	20	AAK72989
23	57	28.9	216	15	AAK63461
24	56.8	28.8	177	16	AAK09148
25	56.8	28.8	177	17	AAK06305
26	56.4	28.6	90	20	AAK06760
27	56.4	28.6	90	22	AAK12812
28	56.4	28.6	90	22	AAK55702
29	56	28.4	213	15	AAK63462
30	52	26.4	772	10	AAK91019
31	51.6	26.2	110	22	AAK12810
32	49.8	25.3	751	20	AAK80793
33	49.8	25.3	771	11	AAK06181
34	49.8	25.3	771	20	AAK62806
35	49.8	25.3	771	20	AAK72998
36	48.4	24.6	91	20	AAK80795
37	48.4	24.6	102	22	AAK12811
38	48.4	24.6	207	21	AAK28519
39	48.4	24.6	242	11	AAK06552
40	47.2	24.0	656	11	AAK06182
41	47.2	24.0	656	20	AAK62807
42	47.2	24.0	656	20	AAK72999
43	43.2	21.9	735	10	AAK91020
44	42.6	21.6	169	19	AAK50383
45	42.4	21.5	123219	23	AAK80703

ALIGNMENTS

RESULT 1
ID AAK59774 standard; cDNA: 197 BP.
AAK59774:
26-JUL-1999 (first entry)
cDNA encoding a thanatone fusion product of the invention.
KW Thanatone; fungal disease; bacterial disease; Cercospora beticola;
KW Cladosporium herbarum; Fusarium cuneiform; F. graminearum;
KW Phytophthora cinamoni; selection marker; plant transformation;
herbicide resistance; PR-1a gene; tobacco; ss.
OS Synthetic.
XX
XX
PN FR2770853-A1.
PD 14-MAY-1999.
PF 07-NOV-1997; 97FR-0014263.
PR 07-NOV-1997; 97FR-0014263.
PR (RHON) RHONE-FOULENC AGROCHIMIE.
PI Derose R. Freyssinet G, Hoffmann J;
XX WPI: 1999-315645/27.
XX P-PSDB; AAY15466.

cDNA encoding a th
Chimeric avirulent
C. fulvum avirulen
PR-1a leader sequen
PR-1a cDNA cloned
Tobacco PR-1a gene
PR-1a protein enco
Promoter of the in
Genomic sequence o
Fragment of tobacc
Fragment of tobacc
Tobacco PR-1a gene
Tobacco PR-1a gene
SPRI-Shiva I fusio
Chimeric Avr9 gene
C-terminus of PR-1
Nicotiana tabacum
Tobacco PR-1a sign
SPRI-SB37 fusion p
Nicotiana tabacum
Nicotiana tabacum
Synthetic ORF5 gen
PR-1b cDNA cloned
Tobacco PR-1b gene
PR-1b protein enco
Primer-2 for const
Nicotiana tabacum
PCR11 DNA coding s
sal10 fused to PRL
PR-1c cDNA cloned
Tobacco PR-1c gene
PR-1c protein enco
Nicotiana tabacum
Linked TMV leader
Human DNA sequence

PT	New nucleic acid encoding thanatine useful as a selection marker for transformation of plants	XX
PS	Claim 8; Page 14-15; 24pp; French.	XX
CC	The specification describes a nucleic acid sequence containing the sequence that encodes thanatine. Plants transformed with this nucleic acid sequence are resistant to fungal and bacterial diseases, specifically those caused by <i>Cercospora beticola</i> , <i>Cladosporium herbarum</i> , <i>Fusarium curvum</i> , <i>F. graminearum</i> and <i>Phytophthora cinamomi</i> . The nucleic acid sequence may also be used as a selection marker for transformation of plants with other coding sequences, e.g. those that impart resistance to herbicides. The present sequence encodes a thanatine fusion product of the invention, comprising the signal peptide of the PR-1a gene of tobacco.	XX
SO	Sequence 197 BP; 32 A; 56 C; 49 G; 60 T; 0 other:	XX
Query Match	100.0%; Score 197; DB 20; Length 197;	
Best Local Similarity	100.0%; Pred. No. 6,4e-51;	
Matches 197; Conservative	0; Mismatches 0; Indels 0; Gaps	
QY	1 ggcgtgacgcacatgggtctgcgtcttctctcagcttccatcttctcttctgttcta 60 	
Db	1 ggcgtgacgcacatgggtctgcgtcttctctcagcttccatcttctcttctgttcta 60	
QY	61 ctctctctcttctcttctgtatctctcaactcttgcgtgcggtttccaaagacgaatgc 120 	
Db	61 ctctctctcttctcttctgtatctctcaactcttgcgtgcggtttccaaagacgaatgc 120	
QY	121 caatcatctactgcacacagagagactggttaagtcgacagagatgttgatcgtcgagcg 180 	
Db	121 caatcatctactgcacacagagagactggttaagtcgacagagatgttgatcgtcgagcg 180	
QY	181 aacgtctcagcagatcc 197 	
Db	181 aacgtctcagcagatcc 197	
RESULT 2		
AAK29729		
ID	AAK29729 standard; DNA; 211 BP.	
XX	AAK29729;	
AC		
XX	22-JUN-1999 (first entry)	
DE	Chimeric signal peptide toxin gene.	
XX		
KW	Toxin; androctonin; scorpion; fusion protein; transgenic plant; resistance; fungus; bacterium; infection; ss.	
XX	Chimeric - Androctonus australis.	
OS	Chimeric - Nicotiana sp.	
XX		
PN	W09909189-A1.	
PD	25-FEB-1999.	
XX		
PE	18-AUG-1998; 98WO-FR01814.	
XX		
PR	20-AUG-1997; 97FR-0010632.	
XX		
PA	(RHON) RHONE-POULENC AGROCHIMIE.	
PI	Derose R, Freyssinet G, Hoffmann J;	
XX		
DR	WPI: 1999-181046/15.	
DR	P-PSDB; W.	
XX		
PT	DNA encoding scorpion peptide androctonin - especially for producing disease-resistant plants	

XX	Claim 17; Page 26; 37pp; French.
PS	
XX	
CC	This sequence corresponds to the coding sequence for a fusion gene
CC	comprising the tobacco PR-1alpha gene signal peptide sequence linked
CC	to the gene encoding the toxin androctonin from the scorpion Androctonus
CC	australis, for expression in plants. Transgenic plants containing
CC	androctonin genes are stated to be resistant to fungal and bacterial
CC	infections, especially caused by Cercospora beticola, Cladosporium
CC	herbarum, Fusarium culmorum, Fusarium graminearum or Phytophthora
CC	cinamomi.
XX	
SO	Sequence 211 BP; 33 A; 58 C; 57 G; 63 T; 0 other;
XX	
Query Match	59.6%; Score 117.4; DB 20; Length 211;
Best Local Similarity	77.0%; Pred. No. 1.7e-26;
Matches 161; Conservative 0; Mismatches 36; Indels 12; Gaps	1;
QY	1 ggcgtgagcgcatgggtctgctggtcttcctcccaactccatcttctcttggtcta 60
Db	1 ggcgtgagcgcatgggtctgctggtcttcctcccaactccatcttctcttggtcta 60
QY	61 ctctctctctcttctcttgatctctcaactctgctgcgtgccaagaagccagtgc 120
Db	61 ctctctctctcttctcttgatctctcaactctgctgcgtgccaagaagccagtgc 120
QY	121 caatcatctactgcacaagagagactgtaagtgcacagga-----gttag 168
Db	121 tcaagatctgcacggagggagggtggtctactacaaatgacactaacagcgcatctgag 180
QY	169 ctgcgcgagcgcaacgctgtcagcagatcc 197
Db	181 ctgcgcgagcgcaacgctgtcagcagatcc 209
XX	
RESULT	3
AAK59773	
ID	AAK59773 standard; cDNA: 106 BP.
XX	
AC	AAK59773;
XX	
DT	26-JUL-1999 (first entry)
XX	
DE	cDNA encoding a thanatine fusion product of the invention.
XX	
XX	Thianatine; fungal disease; bacterial disease; Cercospora beticola;
KW	Cladosporium herbarum; Fusarium culmorum; F. graminearum;
KW	Phytophthora cinamomi; selection marker; plant transformation;
XX	herbicide resistance; PR-1a gene; tobacco; ss.
OS	Synthetic.
XX	
PN	FR2770853-A1.
XX	
PD	14-MAY-1999.
XX	
PF	07-NOV-1997; 97FR-0014263.
XX	
PR	07-NOV-1997; 97FR-0014263.
XX	
PA	(RHON) RHONE-POULENC AGROCHIMIE.
XX	
PI	Derose R, Freyssinet G, Hoffmann J;
XX	
DR	WPI; 1999-315645/27.
XX	
DR	P-PSDB; AAY15465.
XX	
PT	New nucleic acid encoding thanatine useful as a selection marker for
XX	transformation of plants
XX	
XX	Example 1; Page 14; 24pp; French.

CC The specification describes a nucleic acid sequence containing the
 CC sequence that encodes thanatine. Plants transformed with this nucleic
 CC acid sequence are resistant to fungal and bacterial diseases,
 CC specifically those caused by *Cercospora beticola*, *Cladosporium*
 CC *herbarum*, *Fusarium culmorum*, *F. graminearum* and *Phytophthora*
 CC *cinamomi*. The nucleic acid sequence may also be used as a selection
 CC marker for transformation of plants with other coding sequences,
 CC e.g. those that impart resistance to herbicides. The present sequence
 CC encodes a thanatine fusion product of the invention, comprising the
 CC signal peptide of the PR-1a gene of tobacco.

XX Sequence 106 BP; 8 A; 34 C; 20 G; 44 T; 0 other;

Query Match 52.3%; Score 103; DB 20; Length 106;
 Best Local Similarity 100.0%; Pred. No. 3.6e-22;
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagccatgggttcgtgcttcttcacagcttcacatcttctcttgctcta 60
 DB 1 ggcgcagccatgggttcgtgcttcttcacagcttcacatcttctcttgctcta 60
 QY 61 ctctctcttcttcgtgatctcctcactcttcgcgtgcgg 103
 DB 61 ctctctcttcttcgtgatctcctcactcttcgcgtgcgg 103

RESULT 4

AAZ29728 AAZ29728 standard; DNA; 106 BP.

AC AAZ29728;

DT 22-JUN-1999 (first entry)

XX Tobacco PR-1alpha signal peptide sequence.

XX Toxin; androctonin; scorpion; fusion protein; transgenic plant;
 KM resistance; fungus; bacterium; infection; ss.

XX *Nicotiana* sp.

XX MO9909189-A1.

PD 25-FEB-1999.

PF 18-AUG-1998; 98WO-FR01814.

PR 20-AUG-1997; 97FR-0010632.

XX (RHON) RHONE-POULENC AGROCHIMIE.

PI Derose R, Freyssinet G, Hoffmann J;

DR WPI; 1999-181046/15.

XX P-PSDB; AAW99576.

PT DNA encoding scorpion peptide androctonin - especially for producing
 PT disease-resistant plants

PS Claim 15; Page 26; 37pp; French.

XX This sequence corresponds to the signal peptide sequence from the
 CC PR-1alpha gene of tobacco. The sequence is used to generate a fusion
 CC gene with the scorpion androctonin toxin gene from *Androctonus australis*,
 CC for expression in a plant. Transgenic plants containing androctonin
 CC genes are stated to be resistant to fungal and bacterial infections,
 CC especially caused by *Cercospora beticola*, *Cladosporium herbarum*, *Fusarium*
 CC *culmorum*, *Fusarium graminearum* or *Phytophthora cinamomi*.

XX Sequence 106 BP; 8 A; 34 C; 20 G; 44 T; 0 other;

Query Match 52.3%; Score 103; DB 20; Length 106;
 Best Local Similarity 100.0%; Pred. No. 3.6e-22;
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagccatgggttcgtgcttcttcacagcttcacatcttctcttgctcta 60
 DB 1 ggcgcagccatgggttcgtgcttcttcacagcttcacatcttctcttgctcta 60
 QY 61 ctctctcttcttcgtgatctcctcactcttcgcgtgcgg 103
 DB 61 ctctctcttcttcgtgatctcctcactcttcgcgtgcgg 103

RESULT 5

AAZ45694 AAZ45694 standard; DNA; 120 BP.

AC AAZ45694;

DT 06-APR-2000 (first entry)

XX DNA encoding the tobacco PR-1alpha signal peptide.

XX PR-1alpha; heliomyacin; polycysteine peptide; antifungal; antibacterial;
 KM fungal infection; plant protection; animal protection; transgenic plant;
 KM fungi; bacteria; *Cercospora beticola*; *Cladosporium herbarum*;
 KM *Fusarium culmorum*; *F. oxysporum*; *Phytophthora cinamomi*; ss.

XX *Nicotiana* sp.

XX Key Location/Qualifiers

FT CDS 12..102

FT /*tag= a /note= "no termination codon given"

XX FR2777568-A1.

PD 22-OCT-1999.

PF 15-APR-1998; 98FR-0004933.

PR 15-APR-1998; 98FR-0004933.

XX (RHON) RHONE-POULENC AGROCHIMIE.

PI WPI; 2000-108532/10.

DR P-PSDB; AAY54377.

XX New polycysteine peptides, designated heliomyacin, with antifungal or
 PT antibacterial activity, useful in medicine or plant protection -

PS Example 4; Page 33; 46pp; French.

XX The present sequence encodes the tobacco PR-1alpha signal peptide.
 CC This signal sequence is used to produce a fusion peptide with
 CC heliomyacin. The attachment of the PR-1alpha sequence to the N-terminal
 CC aids secretion of the heliomyacin peptide. Heliomyacin is a polycysteine
 CC peptide with antifungal and antibacterial activity, isolated from the
 CC lepidopteran *Heliothis virescens*. Heliomyacin peptides contain the
 CC sequence given in AAY54378. The heliomyacin peptides are used as
 CC pharmaceuticals for treating or preventing fungal infections in humans
 CC and animals. They may also be used as antifungal agents for plant
 CC protection. Nucleic acid encoding the heliomyacin peptides is used to
 CC generate transgenic plants that are resistant to some fungi and
 CC bacteria (specifically *Cercospora beticola*, *Cladosporium herbarum*,
 CC *Fusarium culmorum*, *F. oxysporum* and *Phytophthora cinamomi*), and for
 CC recombinant production of the peptides.

XX Sequence 120 BP; 14 A; 36 C; 23 G; 47 T; 0 other;

Query Match 51.6%; Score 101.6; DB 21; Length 120;
 Best Local Similarity 92.2%; Pred. No. 1e-21;

Matches 107; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 1 ggcgcagacgcatgggttcgtcttctctcagctccatcttctctctgtctta 60
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 1 ggcgcagacgcatgggttcgtcttctctcagctccatcttctctctgtctta 60
 OY 61 ct 116
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 61 ct 116

RESULT 6

AA59772
 ID AA59772 standard; CDNA; 98 BP.

AC AA59772;

DT 26-JUL-1999 (first entry)

DE cDNA encoding the signal peptide of the PR-1a gene of tobacco.

XX Thanatine; fungal disease; bacterial disease; Cercospora beticola;

KW Cladosporium herbarum; Fusarium culmorum; F. graminearum;

KM Phytophthora cinamomi; selection marker; plant transformation;

KM herbicide resistance; PR-1a gene; tobacco; ss.

OS Nicotiana tabacum.

XX FR2770853-A1.

PN 14-MAY-1999.

XX 07-NOV-1997; 97FR-0014263.

XX 07-NOV-1997; 97FR-0014263.

XX (RHON) RHONE-POULENC AGROCHIMIE.

XX Derose R, Freyssinet G, Hoffmann J;

XX WPI; 1999-315645/27.

XX P-PSDB; AAY15464.

XX New nucleic acid encoding thanatine useful as a selection marker for

XX transformation of plants

XX Example 1; Page 14; 24pp; French.

XX The specification describes a nucleic acid sequence containing the

XX sequence that encodes thanatine. Plants transformed with this nucleic

XX acid sequence are resistant to fungal and bacterial diseases,

XX specifically those caused by Cercospora beticola, Cladosporium

XX herbarum, Fusarium culmorum, F. graminearum and Phytophthora

XX cinamomi. The nucleic acid sequence may also be used as a selection

XX marker for transformation of plants with other coding sequences,

XX e.g. those that impart resistance to herbicides. The present sequence

XX encodes the signal peptide of the PR-1a gene of tobacco, which was

XX used to make fusion products of the invention.

CC Sequence 98 BP; 25 A; 23 C; 34 G; 16 T; 0 other;

CC Query Match 48.7%; Score 96; DB 20; Length 98;

CC Best Local Similarity 100.0%; Pred. No. 4.9e-20; Mismatches 0; Indels 0; Gaps 0;

CC Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 ggttcacaagaagcagtcgcaatcatctactgcaacagagagactgtaagtcacagag 161

Db 1 ggttcacaagaagcagtcgcaatcatctactgcaacagagagactgtaagtcacagag 60

OY 162 atgtgaagtcggcagagcagtcgcaacgagatcc 197

Db 61 atgtgaagtcggcagagcagtcgcaacgagatcc 96

RESULT 7

AA245693
 ID AA245693 standard; DNA; 261 BP.

XX AA245693;

DT 06-APR-2000 (first entry)

DE DNA encoding a fusion peptide of PR-1alpha signal peptide/hellomycin.

XX PR-1alpha; hellomycin; polycysteine peptide; antifungal; antibacterial;

KW fungal infection; plant protection; animal protection; transgenic plant;

KM fungi; bacteria; Cercospora beticola; Cladosporium herbarum;

KW Fusarium culmorum; F. oxysporum; Phytophthora cinamomi; ss.

XX Chimeric - Nicotiana sp.

OS Chimeric - Heliothis virescens.

XX Key

FT CDS

FT sig-peptide

FT mat-peptide

FT FR2777568-A1.

XX 22-OCT-1999.

XX 15-APR-1998; 98FR-0004933.

XX 15-APR-1998; 98FR-0004933.

XX (RHON) RHONE-POULENC AGROCHIMIE.

XX WPI; 2000-108532/10.

XX P-PSDB; AAY54376.

XX New polycysteine peptides, designated hellomycin, with antifungal or

XX antibacterial activity, useful in medicine or plant protection -

XX Example 4; Page 32; 46pp; French.

XX The present sequence encodes a fusion peptide of the tobacco

XX PR-1alpha signal peptide and hellomycin. The attachment of the

XX PR-1alpha sequence to the N-terminal aids secretion of the hellomycin

XX peptide. Hellomycin is a polycysteine peptide with antifungal and

XX antibacterial activity, isolated from the lepidopteran Heliothis

XX virescens. Hellomycin peptides contain the sequence given in

XX AAY54378. The hellomycin peptides are used as pharmaceuticals for

XX treating or preventing fungal infections in humans and animals. They

XX may also be used as antifungal agents for plant protection. Nucleic

XX acid encoding the hellomycin peptides is used to generate transgenic

XX plants that are resistant to some fungi and bacteria (specifically

XX Cercospora beticola, Cladosporium herbarum, Fusarium culmorum,

XX F. oxysporum and Phytophthora cinamomi), and for recombinant production

CC Sequence 261 BP; 39 A; 66 C; 74 G; 82 T; 0 other;

CC Query Match 47.4%; Score 93.4; DB 21; Length 261;

CC Best Local Similarity 98.9%; Pred. No. 4.3e-19; Mismatches 1; Indels 0; Gaps 0;

CC Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 10 ccaagggttcggtcttctctcagcttcacatctctctctctgttctactctcttc 69

Db 10 ccaagggttcggtcttctctcagcttcacatctctctctctgttctactctcttc 69

Db 1 ccatgggttcgtgtcttcctcagctccatcttcctctgtgtctactctcttc 60
 Oy 70 ttctcctgtatctctcactctgtccgtcgcgt 104
 Db 61 ttctcctgtatctcctcactctgtccgtcgcgt 95

RESULT 8

AAK29730 ID AAK29730 standard; DNA; 75 BP.

AAK29730:

22-JUN-1999 (first entry)

Oligo #1 for scorpion toxin fusion gene.

XX Toxin; androctonin; scorpion; fusion protein; transgenic plant;

KW resistance; fungus; bacterium; infection; ss.

OS Synthetic.

MO9909189-A1.

25-FEB-1999.

18-AUG-1998; 98WO-FR01814.

20-AUG-1997; 97FR-0010632.

(RHON) RHONE-POULENC AGROCHIMIE.

Derose R, Freyssinet G, Hoffmann J;

WPI; 1999-181046/15.

DNA encoding scorpion peptide androctonin - especially for producing

PT disease-resistant plants

Example 1; Page 11; 37pp; French.

CC This sequence corresponds to an oligonucleotide used to generate a
 CC fusion gene (AAK29729) comprising the tobacco PR-1alpha gene signal
 CC peptide sequence linked to the gene encoding the toxin androctonin
 CC from the scorpion Androctonus australis, for expression in plants.
 CC Transgenic plants containing androctonin genes are stated to be
 CC resistant to fungal and bacterial infections, especially caused by
 CC Cercospora beticola, Cladosporium herbarum, Fusarium culmorum,
 CC Fusarium graminearum or Phytophthora cinamomi.
 SQ Sequence 75 BP; 5 A; 23 C; 13 G; 34 T; 0 other;

Query Match

Best Local Similarity 37.3%; Score 73.4; DB 20; Length 75;
 Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ggcgtgacgcacatgggttcgtcttcctcagctccatcttcctctctgtgcta 60
 Db 1 ggcgtgacgcacatgggttcgtcttcctcagctccatcttcctctctgtgcta 60
 Oy 61 cctctctctcttc 75
 Db 61 cctctctctcttc 75

RESULT 9

AAK29731/C ID AAK29731 standard; DNA; 72 BP.

AAK29731:

22-JUN-1999 (first entry)

XX Oligo #2 for scorpion toxin fusion gene.
 DE Toxin; androctonin; scorpion; fusion protein; transgenic plant;
 KW resistance; fungus; bacterium; infection; ss.

OS Synthetic.

MO9909189-A1.

25-FEB-1999.

18-AUG-1998; 98WO-FR01814.

20-AUG-1997; 97FR-0010632.

(RHON) RHONE-POULENC AGROCHIMIE.

Derose R, Freyssinet G, Hoffmann J;

WPI; 1999-181046/15.

DNA encoding scorpion peptide androctonin - especially for producing

PT disease-resistant plants

Example 1; Page 11; 37pp; French.

CC This sequence corresponds to an oligonucleotide used to generate a
 CC fusion gene (AAK29729) comprising the tobacco PR-1alpha gene signal
 CC peptide sequence linked to the gene encoding the toxin androctonin
 CC from the scorpion Androctonus australis, for expression in plants.
 CC Transgenic plants containing androctonin genes are stated to be
 CC resistant to fungal and bacterial infections, especially caused by
 CC Cercospora beticola, Cladosporium herbarum, Fusarium culmorum,
 CC Fusarium graminearum or Phytophthora cinamomi.
 SQ Sequence 72 BP; 33 A; 11 C; 23 G; 5 T; 0 other;

Query Match

Best Local Similarity 34.2%; Score 67.4; DB 20; Length 72;
 Matches 68; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 35 gcttcacatcttcctctgtgtcactctcttcctctgtgtcactcttg 94
 Db 72 gcttcacatcttcctctgtgtcactctcttcctctgtgtcactcttg 13
 Oy 95 ccgtgcccgg 103
 Db 12 CCGTGCCGG 4

RESULT 10

AAK59771 ID AAK59771 standard; CDNA; 63 BP.

AAK59771:

26-JUL-1999 (first entry)

CDNA encoding a thanatine peptide.

KW Thanatine; fungal disease; bacterial disease; Cercospora beticola;
 KW Cladosporium herbarum; Fusarium culmorum; F. graminearum;
 KW Phytophthora cinamomi; selection marker; plant transformation;
 KW herbicide resistance; ss.

Psodius sp.

FR2770853-A1.

14-MAY-1999.

```

PF 07-NOV-1997; 97FR-0014263.
XX
XX 07-NOV-1997; 97FR-0014263.
XX
PA (RHON ) RHONE-POULENC AGROCHIMIE.
XX
PI Derose R, Freyssinet G, Hoffmann J;
XX
DR WP1: 1999-315645/27.
XX
DR P-PSDB: AAY15463.
XX
PT New nucleic acid encoding thanatine useful as a selection marker for
XX transformation of plants.
XX
PS Claim 4; Page 13; 24pp; French.
XX
CC The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by Cercospora beticola, Cladosporium
CC herbarum, Fusarium culmorum, F. graminearum and Phytophthora
CC cinamomi. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the peptide thanatine.
XX
SQ Sequence 63 BP; 19 A; 14 C; 19 G; 11 T; 0 other;

Query Match 32.08; Score 63; DB 20; Length 63;
Best Local Similarity 100.0%; Pred. No. 5.6e-10;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 ggtccaaagccagtcgcaatcatctactgcaacagagagctgtaagtccagag 161
DB 1 ggtccaaagccagtcgcaatcatctactgcaacagagagctgtaagtccagag 60
OY 162 atg 164
DB 61 atg 63

RESULT 11
AAT06310
ID AAT06310 standard; DNA; 484 BP.
XX
AC AAT06310;
XX
DT 14-APR-1996 (first entry)
XX
DE Chimeric avirulence gene, Avr9.
XX
XX Pathogen resistance; Cf-9; tomato; C.fulvum; Avr 4; Avr 9; fungal;
KW leaf mould; variegation; ds.
XX
XX Lycopersicon esculentum.
OS Cladosporium fulvum.
XX
FH Key Location/Qualifiers
FT CDS 4..415
FT /tag= a
FT /product= secreted_Avr9_protein
XX
XX MO9531564-A2.
XX
XX 23-NOV-1995.
XX
XX 11-MAY-1995; 95MO-GB01075.
XX
XX 07-APR-1995; 95GB-0007232.
XX
XX 11-MAY-1994; 94GB-0009394.
XX
XX 23-DEC-1994; 94MO-GB02812.
XX
XX 31-MAR-1995; 95GB-0006658.

```

```

XX
XX (GATS-) GATSBY CHARITABLE FOUND.
XX
XX Hammond-Kosack KE, Jones DA, Jones JDG;
XX
XX WP1: 1996-010949/01.
XX
XX P-PSDB: AAR85302.
XX
XX Increasing plant pathogen resistance by induction of variegation
XX may lead to acquired resistance to a broad range of pathogens.
XX
XX Claim 7; Page 93; 131pp; English.
XX
XX AAT06310 is the chimeric avirulence (Avr) gene Avr9. It is engineered
XX from the tomato signal peptide sequence of the Pr-1a gene (AAT06305) and
XX the Avr9 gene of C.fulvum. In a new method pathogen resistance genes are
XX expressed highly in genetic constructs which may be used to impart a
XX broad range of pathogen resistance, by induction of variegation, to
XX transgenic plants (or parts or propagules of plants) containing such
XX constructs. The plant pathogen resistance gene Cf-9 imparts resistance
XX to the disease caused by the leaf mould fungal pathogen Cladosporium
XX fulvum. C.fulvum contains avirulence (Avr) genes that confer recognition
XX by plants containing Cf-genes, leading to the activation of host defence
XX mechanisms to attack the disease.
XX
SQ Sequence 484 BP; 128 A; 124 C; 121 G; 111 T; 0 other;

Query Match 31.28; Score 61.4; DB 17; Length 484;
Best Local Similarity 67.7%; Pred. No. 3.4e-09;
Matches 86; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 10 ccatgggttcgtgtcttctctcagcttcacatcttctctctgtgtactctctc 69
DB 3 cgatggagattgtctctcttccacaattgcttcattctctctgtctctacattctct 62
OY 70 ttctccttgatctctcactcttgcgttcggttcacaaagccagtcgaatcatct 129
DB 63 tattcctagtaataaccactcttgcgcgcgaagcccaacaactcaacataaac 122
OY 130 actgcaa 136
DB 123 catgcaa 129

RESULT 12
AAT47876
ID AAT47876 standard; DNA; 484 BP.
XX
AC AAT47876;
XX
DT 27-MAR-1997 (first entry)
XX
DE C. fulvum avirulence AVR4 fusion gene.
XX
XX Pathogen resistance; disease resistance; AVR4 gene; avirulence;
KW tomato leaf mould; Lycopersicon; Cf-4 gene; transgenic plant;
XX crop protection; elicitor; ds.
XX
XX Chimeric Nicotiana sp.
OS Chimeric Cladosporium fulvum.
XX
FH Key Location/Qualifiers
FT CDS 5..415
FT /tag= a
FT sig_peptide 5..94
FT /tag= b
FT /note= "tobacco Pr1a protein signal sequence"
FT mat_peptide 95..412
FT /tag= c
FT /note= "C. fulvum Avr4"
XX
XX MO9635790-A1.

```



```
FT      /*tag- a
FT      /product=PR-1a
XX      EP392225-A.
XX      17-OCT-1990.
XX      21-MAR-1990; 90EP-0105336.
XX      20-OCT-1989; 89US-0425504.
XX      24-MAR-1989; 89US-0329018.
XX      20-JUN-1989; 89US-0368672.
XX      (CIBA ) CIBA GEIGY AG.
XX      Ryals JA, Alexander DC, Goodman RM, Melns F, Payne GB;
XX      Stinson JR, Neuhaus J-M, Moyer MB;
XX      WPI: 1990-313983/42.
XX      P-PSDB: AAR07313.
XX      Disease-resistant transgenic plants - obtd. using encoding an
XX      Inductible pathogenesis-related protein from infected plants.
XX      Example 15; page 23; 77pp: English.
XX      This chimeric DNA sequence comprises the PR-1a plant pathogenesis-
XX      related protein (PRP)-coding sequence (I), and a promoter sequence
XX      which enhances transcription of (I). This construct is used to
XX      produce transgenic plant cells or -tissues with the ability to re-
XX      generate into plants which are disease resistant.
XX      See also AA006179, AA006181-86, AA006199-Q06208 and AA006829.
XX      Sequence 809 BP; 231 A; 153 C; 171 G; 254 T; 0 other:
XX
XX      Query Match      29.3%; Score 57.8; DB 11; Length 809;
XX      Best Local Similarity 76.3%; Pred. No. 5.2e-08;
XX      Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
XX
XX      QY      9 gccatggattcgctctctcagctccatctctcctctgtgctactctctt 68
XX      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      Db      27 gtcattggattgtctctcttccaatgctcctcatctctctgtctcaactcttc 86
XX
XX      QY      69 ctttccctgtgactcctcaactctgcccgtgcc 101
XX      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      Db      87 ttattcctagtaataatcccaactctgcccgtgcc 119
XX
XX      RESULT 15
XX      ID      AAV62305
XX      AAV62805 standard; CDNA; 809 BP.
XX
XX      AC      AAV62805;
XX      DT      05-MAR-1999 (first entry)
XX      XX
XX      DE      Tobacco PR-1a gene clone.
XX      XX
XX      KM      Chemically regulatable DNA promoter: expression control; pesticide;
XX      herbicide tolerance; pathogenesis related gene; PR gene; ss.
XX      OS      Nicotiana acuminata.
XX      OS      US5851766-A.
XX      PD      22-DEC-1998.
XX      PE      31-MAY-1995; 95US-0456262.
XX      PR      31-MAY-1995; 95US-0456262.
XX      PA      (NOVS ) NOVARTIS FINANCE CORP.
```

```
XX      Harms C, Ryals JA;
XX      WPI: 1999-080396/07.
XX      Isolating chemically regulatable DNA sequences in plants - useful
XX      for chemically controlling expression in transformed plants
XX      Example 44; Column 179-182; 175pp: English.
XX
XX      This sequence represents a clone of the tobacco pathogenesis related (PR)
XX      gene. This gene can be isolated using the method of the invention.
XX      The method is for isolating a chemically regulatable DNA promoter
XX      fragment from the 5' flanking region of a chemically regulatable gene in
XX      a plant tissue. The method allows isolation of sequences which will be
XX      useful for the controlled expression of genes, under the control of a
XX      non-coding regulatable sequence. This is useful in plants with a
XX      herbicide or pesticide detoxification mechanism under the control of a
XX      chemical regulator, the regulator being applied before or with the
XX      herbicide or pesticide to give optimal tolerance. The promoter fragment
XX      is useful for controlling sequences which encode traits such as
XX      height, shape, development, male or female sterility, and the ability
XX      of the plant to withstand cold, heat, salt and drought. The chemical
XX      induction of the promoter allows the regulation of production of
XX      compounds, e.g. flavours, fragrances, pigments, natural sweeteners,
XX      industrial feedstocks, antimicrobials and pharmaceuticals, by
XX      biosynthesis or metabolic conversion, whose biosynthesis is controlled
XX      by endogenous or foreign genes. The method allows control over the time
XX      and rate of gene expression either throughout the whole plant, or in
XX      localized tissues, to achieve e.g. fungal or insect resistance by for
XX      instance dusting the leaves with the chemical regulator. Controlling the
XX      developmental processes by the application of a regulating chemical in
XX      e.g. the commercial production of cultivated crops allows processes such
XX      as germination, flower formation and fruit ripening to be synchronised at
XX      a given time.
XX
XX      Sequence 809 BP; 231 A; 153 C; 171 G; 254 T; 0 other:
XX
XX      Query Match      29.3%; Score 57.8; DB 20; Length 809;
XX      Best Local Similarity 76.3%; Pred. No. 5.2e-08;
XX      Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
XX
XX      QY      9 gccatggattcgctctctcagctccatctctcctctgtgctactctctt 68
XX      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      Db      27 gtcattggattgtctctcttccaatgctcctcatctctctgtctcaactcttc 86
XX
XX      QY      69 ctttccctgtgactcctcaactctgcccgtgcc 101
XX      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      Db      87 ttattcctagtaataatcccaactctgcccgtgcc 119
XX
```

Search completed: May 12, 2002, 20:20:40
Job time: 4927 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 20:14:35 ; Search time 76.1 Seconds
(without alignments)
203.350 Million cell updates/sec

Title: US-09-554-024-2
Perfect score: 63
Sequence: 1 ggtccagaagccagtcgccc.....ctgtaagtcgacagagatg 63

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_NA: *
1: /cgn2_6/prodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/prodata/2/ina/5A.COMB.seq: *
4: /cgn2_6/prodata/2/ina/5B.COMB.seq: *
5: /cgn2_6/prodata/2/ina/5A.COMB.seq: *
6: /cgn2_6/prodata/2/ina/5B.COMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	23.4	37.1	3102	1	US-08-462-484-5
2	23.4	37.1	3102	1	US-08-441-147-5
3	23.4	37.1	3102	5	PCT-US95-07536-5
4	22.8	36.2	80	1	US-08-418-859-54
5	22.8	36.2	80	1	US-08-418-859-56
6	22.8	36.2	80	2	US-08-643-181-54
7	22.8	36.2	80	2	US-08-643-181-56
8	22.8	36.2	80	2	US-08-643-181-54
9	21.8	34.6	180	5	PCT-US93-03076-1
10	21.6	34.3	330	5	US-08-358-160-145
11	21.6	34.3	375	5	PCT-US91-02766-13
12	21.6	34.3	395	6	5428135-3
13	21.6	34.3	376	6	US-08-998-416-868
14	21.6	34.3	376	6	5194596-12
15	21.6	34.3	649	6	5219739-12
16	21.6	34.3	748	1	US-08-387-845-1
17	21.6	34.3	748	2	US-08-778-275-1
18	21.6	34.3	748	2	US-08-667-352-1
19	21.6	34.3	1247	5	PCT-US91-02766-19
20	21.6	34.3	1247	5	5219759-3
21	21.6	34.3	1316	5	PCT-US91-02766-21
22	21.2	33.7	80	1	US-08-418-859-55
23	21.2	33.7	80	1	US-08-643-181-55
24	21.2	33.7	1074	2	US-08-627-151A-15
25	21.2	33.7	1404	6	5171840-8
26	21.2	33.7	1404	6	5480796-8
27	21.2	33.7	1486	4	US-08-795-473B-3

28	21.2	33.7	2061	6	5171840-1	Patent No. 5171840
29	21.2	33.7	2061	6	5480796-1	Patent No. 5480796
30	21.2	33.7	3319	4	US-08-795-473B-2	Sequence 2, App1
31	21.2	33.7	3319	4	US-08-358-160-143	Sequence 143, App
32	21.2	33.3	234	1	US-08-253-155A-8	Sequence 8, App1
33	21.2	33.3	638	3	US-08-705-771-11	Sequence 11, App1
34	21.2	33.3	655	21	US-09-109-204-27	Sequence 27, App1
35	21.2	33.3	1882	3	US-08-501-572-4	Sequence 4, App1
36	21.2	33.3	1882	3	US-09-104-444-4	Sequence 4, App1
37	21.2	33.3	2224	4	US-09-109-204-6	Sequence 6, App1
38	20.8	33.0	1773	4	US-08-943-731-215	Sequence 215, App
39	20.8	33.0	1966	6	5475086-3	Patent No. 5475086
40	20.8	33.0	2026	1	US-08-485-718-8	Sequence 8, App1
41	20.8	33.0	2026	2	US-08-484-530-54	Sequence 54, App1
42	20.8	33.0	2026	2	US-08-827-618A-54	Sequence 54, App1
43	20.8	33.0	2026	3	US-08-483-952A-54	Sequence 54, App1
44	20.8	33.0	6246	4	US-08-943-731-640	Sequence 640, App
45	20.6	32.7	699	4	US-08-991-789A-172	Sequence 172, App

ALIGNMENTS

RESULT 1
US-08-462-484-5
Sequence 5, Application US/08462484
Patent No. 5667331
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Xu, Feng
APPLICANT: Dalboge, Henrik
APPLICANT: Schneider, Palle
APPLICANT: Asalyng, Dorrit A.
TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: No. 56675310 No. 5667531disk of No. 5667531th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,484
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/441,147
FILING DATE: 15-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lowmey, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4185.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 878 9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Polyporus plinthus
FEATURE:
NAME/KEY: intron

CORRESPONDENCE ADDRESS:
ADDRESSEE: Novo Nordisk of North America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07536
FILING DATE: 15-June-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/265,534
FILING DATE: 24-June-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4185.204-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 878 9635
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Polyporus pinastus
FEATURE:
NAME/KEY: intron
LOCATION: 666..720
FEATURE:
NAME/KEY: intron
LOCATION: 790..845
FEATURE:
NAME/KEY: intron
LOCATION: 1125..1182
FEATURE:
NAME/KEY: intron
LOCATION: 1390..1450
FEATURE:
NAME/KEY: intron
LOCATION: 1607..1661
FEATURE:
NAME/KEY: intron
LOCATION: 1863..1918
FEATURE:
NAME/KEY: intron
LOCATION: 1976..2025
FEATURE:
NAME/KEY: intron
LOCATION: 2227..2285
FEATURE:
NAME/KEY: intron
LOCATION: 2403..2458
FEATURE:
NAME/KEY: intron
LOCATION: 2576..2627
FEATURE:
NAME/KEY: CDS
LOCATION: join (665..721, 789..846, 1124..1183, 1389..1451,
1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
2575..2628).
PCT-US95-07536-5

Query Match 37.1%; Score 23.4; DB 5; Length 3102;
Best Local Similarity 73.2%; Pred. No. 6.5; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 11;

QY 14 cagtgccaatcatctactgcaacagagagacggtgaagtc 54
Db 2614 CACAGTGACTCATCTTTTGCACAGAGAGACTGACACAGC 2654

RESULT 4
US-08-418-859-54
Sequence 54, Application US/08418859
Patent No. 5811235
GENERAL INFORMATION:
APPLICANT: Jeffreys, Alec J.
TITLE OF INVENTION: METHOD OF
TITLE OF INVENTION: CHARACTERISATION
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cushman, Darby & Cushman
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.2 Mb
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS 3.20
SOFTWARE: ASCII from WPS-PLUS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418, 859
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/935,107
FILING DATE: 26 August 1992
APPLICATION NUMBER: 9118371.5
FILING DATE: 27-Aug-1991
APPLICATION NUMBER: 9119089.2
FILING DATE: 06-Sep-1991
APPLICATION NUMBER: 9124636.3
FILING DATE: 20-No. 5811235-1991
APPLICATION NUMBER: 9207379.0
FILING DATE: 03-Apr-1992
APPLICATION NUMBER: 9212627.5
FILING DATE: 15-Jun-1992
APPLICATION NUMBER: 9212881.8
FILING DATE: 17-Jun-1992
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 97279/PHM.36520/US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (292) 861-3000
TELEFAX: (292) 822-0944
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 80 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-418-859-54

Query Match 36.2%; Score 22.8; DB 1; Length 80;
Best Local Similarity 66.0%; Pred. No. 3.1; Indels 0;
Matches 33; Conservative 0; Mismatches 17;

QY 10 aagccagtgccaatcatctactgcaacagagagacggtgaagtcagag 59
Db 26 AACGTAATCCGATTGGCTACTTAAGAGAGAGAGGATATGAGCAGAG 75

RESULT 5
US-08-418-859-56

```
; Sequence 56, Application US/08418859
; Patent No. 5811235
; GENERAL INFORMATION:
; APPLICANT: Jeffreys, Alec J.
; TITLE OF INVENTION: METHOD OF
; TITLE OF INVENTION: CHARACTERISATION
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.2 Mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS 3.20
; SOFTWARE: ASCII from WPS-PLUS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,859
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/935,107
; FILING DATE: 26 August 1992
; APPLICATION NUMBER: 9118371.5
; FILING DATE: 27-Aug-1991
; APPLICATION NUMBER: 9119089.2
; FILING DATE: 06-Sep-1991
; APPLICATION NUMBER: 9124636.3
; FILING DATE: 20-No. 5811235-1991
; APPLICATION NUMBER: 9207379.0
; FILING DATE: 03-Apr-1992
; APPLICATION NUMBER: 9212627.5
; FILING DATE: 15-Jun-1992
; APPLICATION NUMBER: 9212881.8
; FILING DATE: 17-Jun-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 97279/PHM.36520/US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (292) 861-3000
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-418-859-56

Query Match          36.2%; Score 22.8; DB 1; Length 80;
Best Local Similarity 66.0%; Pred. No. 3.1;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 10 aagccagtgccatcatctactgcacagagactgtaagtgcagag 59
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 26 AAGCTAATTCGATGCTACTTTAAAGAGACAGGCGTATGACGACAG 75

RESULT 6
US-08-643-181-54
; Sequence 54, Application US/08643181
; Patent No. 5853989
; GENERAL INFORMATION:
; APPLICANT: Jeffreys, Alec J.
; TITLE OF INVENTION: METHOD OF
; TITLE OF INVENTION: CHARACTERISATION
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.2 Mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS 3.20
; SOFTWARE: ASCII from WPS-PLUS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/643,181
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,859
; FILING DATE:
; APPLICATION NUMBER: 07/935,107
; FILING DATE: 26 August 1992
; APPLICATION NUMBER: 9118371.5
; FILING DATE: 27-Aug-1991
; APPLICATION NUMBER: 9119089.2
; FILING DATE: 06-Sep-1991
; APPLICATION NUMBER: 9124636.3
; FILING DATE: 20-No. 5853989-1991
; APPLICATION NUMBER: 9207379.0
; FILING DATE: 03-Apr-1992
; APPLICATION NUMBER: 9212627.5
; FILING DATE: 15-Jun-1992
; APPLICATION NUMBER: 9212881.8
; FILING DATE: 17-Jun-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 97279/PHM.36520/US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (292) 861-3000
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-643-181-54

Query Match          36.2%; Score 22.8; DB 2; Length 80;
Best Local Similarity 66.0%; Pred. No. 3.1;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 10 aagccagtgccatcatctactgcacagagactgtaagtgcagag 59
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 26 AAGCTAATTCGATGCTACTTTAAAGAGACAGGCGTATGACGACAG 75

RESULT 7
US-08-643-181-56
; Sequence 56, Application US/08643181
; Patent No. 5853989
; GENERAL INFORMATION:
; APPLICANT: Jeffreys, Alec J.
; TITLE OF INVENTION: METHOD OF
; TITLE OF INVENTION: CHARACTERISATION
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: PCT/US93/03076
3 FILING DATE: 19930331
4 CLASSIFICATION:
5 ATTORNEY/AGENT INFORMATION:
6 NAME: Granahan, Patricia
7 REGISTRATION NUMBER: 32,227
8 REFERENCE/DOCKET NUMBER: WH192-03A
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: 617-861-6240
11 TELEFAX: 617-861-9540
12 INFORMATION FOR SEQ ID NO: 1:
13 SEQUENCE CHARACTERISTICS:
14 . LENGTH: 8298 base pairs
15 . TYPE: NUCLEIC ACID
16 . STRANDEDNESS: double
17 . TOPOLOGY: linear
18 MOLECULE TYPE: DNA (genomic)
19 FEATURE:
20 NAME/KEY: CDS
21 LOCATION: 731..5272
22 PCT-US93-03076-1
23
24 Query Match 36.2% Score 22.8: DB 5: Length 8298:
25 Best Local Similarity 62.1%: Fred. No. 16:
26 Matches 36: Conservative 0: Mismatches 22: Indels 0: Gaps 0:
27
28 QY 1 gttctcaagaagacgagtcgaatctactctgcacagagagagctgtaagtcgcaga 58
29 111111 111111 111111 111111 111111 111111 111111 111111
30 Db 4761 gctttccacacgaagtcattgcattgctgtatgtaccacagggggctctgtacctccgaga 4704
31
32 RESULT 9
33 US-08-358-160-145
34 Sequence 145, Application US/08358160
35 Patent No. 5663143
36 GENERAL INFORMATION:
37 APPLICANT: LEY, Arthur C.
38 APPLICANT: LADNER, Robert C.
39 APPLICANT: GUTERMAN, Sonia K.
40 APPLICANT: ROBERTS, Bruce L.
41 APPLICANT: MARKLAND, William
42 APPLICANT: KENT, Rachel B.
43 TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
44 NUMBER OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
45 NUMBER OF SEQUENCES: 234
46 CORRESPONDENCE ADDRESS:
47 ADDRESSEE: BROWDY AND NEIMARK
48 STREET: 419 Seventh Street, N.W. Suite 300
49 CITY: Washington
50 STATE: District of Columbia
51 COUNTRY: USA
52 ZIP: 20004
53 COMPUTER READABLE FORM:
54 MEDIUM TYPE: Floppy disk
55 OPERATING SYSTEM: PC-DOS/MS-DOS
56 SOFTWARE: PatentIn Release #1.0, Version #1.25
57 CURRENT APPLICATION DATA:
58 APPLICATION NUMBER: US/08/358,160
59 FILING DATE: 16-DEC-1994
60 CLASSIFICATION: 514
61 PRIOR APPLICATION DATA:
62 APPLICATION NUMBER: US 08/133,031
63 FILING DATE: 13-OCT-1993
64 PRIOR APPLICATION DATA:
65 APPLICATION NUMBER: US 08/009,319
66 FILING DATE: 26-JAN-1993
67 PRIOR APPLICATION DATA:
68 APPLICATION NUMBER: US 07/664,989
69 FILING DATE: 01-MAR-1991
70 PRIOR APPLICATION DATA:
71

```

APPLICATION NUMBER: US 07/487,063
 FILING DATE: 02-MAR-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/240,160
 FILING DATE: 02-SEP-1988
 ATTORNEY/AGENT INFORMATION:
 NAME: Cooper, Iyer P
 REGISTRATION NUMBER: 28,005
 REFERENCE/DOCKET NUMBER: Lev-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-628-5197
 TELEFAX: 202-737-3528
 TELEEX: 24863
 INFORMATION FOR SEQ ID NO: 145:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 180 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: synthetic DNA fragment

Query Match	34.6%	Score 21.8;	DB 1;	Length 180;
Best Local Similarity	69.7%;	Pred. No. 10;		
Matches 23; Conservative	5;	Mismatches 5;	Indels 0;	Gaps 0

```

Qy      26 tctactgcacacagagagactcgtgaagtccaga 58
          ::||| ||| ||| |||||:::|:
Db      62 DSTACTACAACTCGAGTATTGGTAACTGCSGv 94

```

RESULT 10
 PCT-US91-02766-13
 Sequence 13, Application PC/Tus9102766
 GENERAL INFORMATION:
 APPLICANT: NASCIMENTO, CARLOS G.
 APPLICANT: CALDERON-CACIA, MARIA D.
 TITLE OF INVENTION: GLYCOSYLATED PDGF
 NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Itrell & Manella
 STREET: 545 Middlefield Road, Suite 200
 City: Menlo Park
 STATE: California
 COUNTRY: USA
 ZIP: 94025
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US91/02766
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/515,474
 FILING DATE: 26-APR-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: ROBINS, ROBERTA L.
 REGISTRATION NUMBER: 33,208
 REFERENCE/DOCKET NUMBER: 2300-0105.40
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 327-7250
 TELEFAX: (415) 327-2951
 TELEX: 706141
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 330 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single

```

; TOPOLOGY: linear
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..330
;
PCT-US91-02766-13

```

Query Match	34.3%;	Score 21.6;	DB 5;	Length 330;
Best Local Similarity	75.0%;	Pred. No. 15;		
Matches 27; Conservative	0;	Mismatches 9;	Indels 0;	Gaps 0;

```

Qy      22 atcatctctgcacagcagagactcgttaagtcgcag 57
          | | | | | | | | | | | | | | | | | |
Db      130 ACCGGCTCTGCACACAGAGCAGTGTCAACATGCCAG 165

```

```

1      RESULT 11
2      PCT-US91-02766-15
3      Sequence 15, Application PC/TUS9102766
4      GENERAL INFORMATION:
5      APPLICANT: NASCIMENTO, CARLOS G.
6      APPLICANT: CALDERON-CACIA, MARIA D.
7      TITLE OF INVENTION: GLYCOSYLATED PDGF
8      NUMBER OF SEQUENCES: 24
9      CORRESPONDENCE ADDRESS:
10     ADDRESSEE: Itrell & Manella
11     STREET: 345 Middlefield Road, Suite 200
12     City: Menlo Park
13     STATE: California
14     COUNTRY: USA
15     ZIP: 94025
16
17     COMPUTER READABLE FORM:
18     MEDIUM TYPE: Floppy disk
19     COMPUTER: IBM PC compatible
20     OPERATING SYSTEM: PC-DOS/MS-DOS
21     SOFTWARE: PatentIn Release #1.0, Version #1.25
22     CURRENT APPLICATION DATA:
23     FILING DATE:
24     CLASSIFICATION:
25     PRIOR APPLICATION DATA:
26     APPLICATION NUMBER: US 07/515,474
27     FILING DATE: 26-APR-1990
28     ATTORNEY/AGENT INFORMATION:
29     NAME: ROBINS, ROBERTA L.
30     REGISTRATION NUMBER: 33,208
31     REFERENCE/DOCKET NUMBER: 2300-0105.40
32     TELECOMMUNICATION INFORMATION:
33     TELEPHONE: (415) 327-7250
34     TELEFAX: (415) 327-2951
35     TELEX: 706141
36     INFORMATION FOR SEQ ID NO: 15:
37     SEQUENCE CHARACTERISTICS:
38     LENGTH: 375 base pairs
39     TYPE: NUCLEIC ACID
40     STRANDEDNESS: single
41     TOPOLOGY: linear
42     FEATURE:
43     NAME/KEY: CDS
44     LOCATION: 1..375
45
46     CDT-US91-02766-15

```

Query Match	34.3%	Score 21.6;	DB 5;	Length 375;
Best Local Similarity	75.0%	Pred. No. 16;		
Matches 27; Conservative	0;	Mismatches 9;	Indels 0;	Gaps 0;

QY 22 atcatctactgcacacaggaagaactgtaagtccag 57
| | | | | | | | | | | | | | | | | |
Db 130 ACCGCGTCTGCACACAGACGACGATGTCAAGTCCAG 165

RESULT 12

	Matches	27;	Conservative	0;	Mismatches	9;	Indels	0;	Gaps	0;
Oy	22	atcatctactgcacacagaggagactggttaagtgccag	57							
Db	396	accggtctgtgcacacgacgacgltgtcaagtlgcag	431							

Search completed: May 12, 2002, 20:14:36
 Job time: 7193 sec

PT New nucleic acid encoding thiamine useful as a selection marker for
XX transformation of plants
PS Claim 4; Page 13; 24pp; French.

CC The specification describes a nucleic acid sequence containing the
CC sequence that encodes thiamine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by *Cercospora blight*, *Gladiosporium*
CC *thamni*, *Fusarium moniliforme*, *F. graminearum* and *Phytophthora*
CC *chlamoni*. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the peptide thiamine.

50 Sequence 63 BP; 19 A; 14 C; 19 G; 11 T; 0 other;

Query Match	100.0%	Score 63:	DB 20:	Length 63:
Best Local Similarity	100.0%	Pred. No	2e-13:	
Matches 63, Conservative	0:	Mismatches	0:	Gaps 0

QY 1 ggtcccaagaaagccagtcgccaatcattctactcgcaacacagagagtgatctgtatagtcgccaagag 600
|||||
Db 1 ggtcccaagaaagccagtcgccaatcattctactcgcaacacagagagtgatctgtatagtcgccaagag 600

Qy	61 atq 63
Db	61 atq 63

RESULT	2
AA59772	
ID	AA59772 standard; cDNA; 98 BP.

AC	AA59772;
XX	
DT	26-JUL-1999 (first entry)

DE cDNA encoding the signal peptide of the PR-1a gene of tobacco.
XX
KM Thanatine; fungal disease; bacterial disease; *Cercospora beticola*

KW Cladosporeum herbarum; Fusarium cuniorum; F. graminearum;
KW Phytophthora cinnamomi; selection marker; plant transformation;
KW herbicide resistance; PR-1a gene; tobacco; ss.

OS	Nicotiana tab
XX	
PN	FR2770853-A1

PD	74-MAY-1999.
XX	
PF	07-NOV-1997: 97FR-0014263.

PR 07-NOV-1997; 97FR-0014263.
XX
PA (RHON) RHONE-POULENC AGROCHIMIE.

PI Derose R, Freyssinet G, Hoffmann J,
XX
DR WPI; 1999-315645/27.

XX New nucleic acid encoding chitinase useful as a selection marker for
PT transformation of plants

Example 1; Page 14; 24pp; French
The specification describes a nu

CC sequence encodes unadorned. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by *Cercospora beticola*, *Cladosporium*
CC *herbarum*, *Fusarium culmorum*, *F. graminearum* and *Phytophthora*

CC cinnamoni. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the signal peptide of the P-1a gene of tobacco, which was
CC used to make fusion products of the invention.

CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the signal peptide of the PR-1a gene of tobacco, which was
CC used to make fusion products of the invention.

Sequence 98 BP; 25 A; 23 C; 34 G; 16 T; 0 other;

Query Match	100.0%;	Score 63;	DB 20;	Length 98;
Best Local Similarity	100.0%;	Pred. No. 2,2e-13;		
Matches 63;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 ggttccaagaagccagtcgccaatcatctacttcagcaacagagagctctgtaagtcgccaagg 600
|||||
Db 1 ggttccaagaagccagtcgccaatcatctacttcagcaacagagagctctgtaagtcgccaagg 600

Qy	61	atg	63
Db	61	atg	63

RESULT	3
AAx59774	

XX
AC
XX
AA59774;
XX

XX cDNA encoding a thanatine
DE
XX

KW Cladosporeum herbarum; Fusarium cunorum; F. graminearum;
KW Phytophthora cinnamomi; selection marker; plant transformation;
KW herbicide resistance; PR-1a gene; tobacco; ss.

OS	Synthetic.
XX	
PN	FR2770853-A1

PD	14-MAY-1999.
XX	
PF	07-NOV-1997; 97FR-0014263.

PR 07-NOV-1997; 97FR-0014263.
XX
PA (RHON) RHONE-POULENC AGROCH

PI Derose R, Freyssinet G, Hoffmann J;
XX
DR WPI; 1999-315645/27.

XX	New nucleic acid en-
PT	transformation of pl
PT	

The specification describes a nuclear reactor core having a central fuel assembly surrounded by a reflector assembly.

CC sequence that encodes chitinase. Plants transformed with this nuclear
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by *Cercospora beticola*, *Cladosporium*
CC *herbarum*, *Fusarium culmorum*, *F. camigeraum* and *Phytophthora*

CC cinnamoni. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes a thannating fusion product of the invention comprising the

CC signal peptide of the PR-1a gene of tobacco.
XX
SQ, sequence 197 BP; 32 A; 56 C; 49 G; 60 T; 0 other

Query Match 100.0%; Score 63; DB 20; Length 197;

Best Local Similarity 100.0%; Pred. No. 2 6e-13; Mismatches 0; Indels 0; Gaps 0;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggltccaagaagcagtgccaatcatctactgcaacagaggaactggtgaagtcacagag 60
 |||
 DB 102 ggltccaagaagcagtgccaatcatctactgcaacagaggaactggtgaagtcacagag 161

OY 61 atg 63
 |||

DB 162 atg 164

RESULT 4

AAK59770
 ID AAK59770 standard; cDNA; 33 BP.

AC AAK59770;

DT *26-JUL-1999 (first entry)

XX CDNA encoding a thanatine peptide.

DE Thanatine: fungal disease; bacterial disease; Cercospora beticola;

KM Cladosporium herbarum; Fusarium culmorum; F. graminearum;

KW Phytophthora cinamomi; selection marker; plant transformation;

KW herbicide resistance; ss.

XX Psodius sp.

XX FR2770853-A1.

XX 14-MAY-1999.

XX 07-NOV-1997; 97FR-0014263.

XX 07-NOV-1997; 97FR-0014263.

PA (RHON) RHONE-POULENC AGROCHIMIE.

PI Deroose R, Freyssinet G, Hoffmann J;

XX WPI: 1999-315645/27.

XX P-PSDB; AAY15462.

PT New nucleic acid encoding thanatine useful as a selection marker for

PT transformation of plants

PS Claim 3; Page 13; 24pp; French.

XX The specification describes a nucleic acid sequence containing the

CC sequence that encodes thanatine. Plants transformed with this nucleic

CC acid sequence are resistant to fungal and bacterial diseases,

CC specifically those caused by Cercospora beticola, Cladosporium

CC herbarum, Fusarium culmorum, F. graminearum and Phytophthora

CC cinamomi. The nucleic acid sequence may also be used as a selection

CC marker for transformation of plants with other coding sequences,

CC e.g. those that impart resistance to herbicides. The present sequence

CC encodes the peptide thanatine.

XX Sequence 33 BP; 10 A; 7 C; 9 G; 7 T; 0 other;

Query Match 52.4%; Score 33; DB 20; Length 33;

Best Local Similarity 100.0%; Pred. No. 0.011; Mismatches 0; Indels 0; Gaps 0;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 atcatctactgcaacagaggaactggtgaagtc 54
 |||
 DB 1 atcatctactgcaacagaggaactggtgaagtc 33

RESULT 5

AAH87754/C
 ID AAH87754 standard; cDNA; 589 BP.

AC AAH87754;

DT 25-SEP-2001 (first entry)

XX Peppermint plant oil gland expressed cDNA 110.

DE Peppermint; plant oil gland cell; terpenoid essential oil; resin;

KW genetic mapping; antisense suppression; recombinant expression; ss.

XX Mentha x piperita.

XX WO200153319-A1.

XX 26-JUL-2001.

XX 19-JAN-2001; 2001WO-US02567.

XX 20-JAN-2000; 2000US-0177264.

XX (CROF/) CROTEAU R B.

XX (LANG/) LANGE B M R.

XX (WILD/) WILDUNG M R.

XX Croteau RB, Lange BM, Wildung MR;

XX WPI: 2001-488706/53.

XX New nucleic acid molecules corresponding to mRNA molecules expressed in

XX peppermint oil glands for enhancing expression of plant oil gland cell

XX proteins -

XX Claim 1; Page 119-120; 251pp; English.

XX The invention relates to nucleic acid molecules (AAH87645-AAH88116) that

XX correspond to all or part of a mRNA molecule expressed in plant oil

XX gland cells, especially peppermint and plant oil glands that produce

XX terpenoid essential oils and resins. The nucleic acids are useful for

XX genetically mapping a plant genome for genes expressed in plant oil

XX gland cells and to suppress (for example by antisense suppression) or

XX enhance their expression (for example by genetically transforming a

XX plant cell with a replicable expression vector that expresses one or more

XX proteins naturally expressed in plant oil gland cells). The nucleic acids

XX are also useful for recombinant expression of plant oil gland proteins

XX required for terpenoid essential oil and/or resin production in bacterial

XX and/or yeast cells.

XX Sequence 589 BP; 164 A; 115 C; 135 G; 175 T; 0 other;

Query Match 42.5%; Score 26.8; DB 22; Length 589;

Best Local Similarity 73.9%; Pred. No. 3.8;

Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 9 gaagccagtcgaatcatctactgcaacagaggaactggtgaagtc 54
 |||
 DB 409 GAGGCCAGTTTATCATCATCAGCAAGAACACAGCACTGGTACACG 364

RESULT 6

AAFL2996/C

ID AAFL2996 standard; cDNA; 2022 BP.

AC AAFL2996;

DT 13-MAR-2001 (first entry)

XX Aspergillus oryzae EST SEQ ID NO:5519.

XX Multiple gene expression; filamentous fungal cell; EST;

XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KM Aspergillus oryzae; Trichoderma reesei; Identification; recombination;
KM culture condition; environmental stress; spore morphogenesis;
KM metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
PI Berka RM, Rey MM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
PS
PS Claim 88; Page 2292-2293; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 2022 BP; 481 A; 597 C; 483 G; 461 T; 0 other;

Query Match 41.6%; Score 26.2; DB 21; Length 2022;
Best Local Similarity 67.3%; Pred. No. 8.5;
Matches 37; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 9 gaaccagtgcaatcctactgcacacagagactggaatgcacagagatg 63
DB 927 GATGCGAGGCTCAATCTTCAGACACATCGAAGAGGAGTTCAGGGGATG 873

RESULT 7
AAAF07491/c
ID AAF07491 standard; cDNA; 1710 BP.
XX
AC AAF07491;
XX
XX 13-MAR-2001 (first entry)
DT
XX
DE Fusarium venenatum EST SEQ ID NO:14.

XX Multiple gene expression; filamentous fungal cell; EST;
KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KM Aspergillus oryzae; Trichoderma reesei; Identification; recombination;
KM culture condition; environmental stress; spore morphogenesis;
KM metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
PI Berka RM, Rey MM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
PS
PS Claim 86; Page 368; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 1710 BP; 395 A; 508 C; 401 G; 399 T; 7 other;

Query Match 39.7%; Score 25; DB 21; Length 1710;
Best Local Similarity 69.4%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 14 cagtgcaatcctactgcacacagagactggaatgcacagagatg 62
DB 624 GAGTCCAAATCTTCATGCGACATGAGAGGGGAAAGTTCAGGGGAT 576

RESULT 8
AAA91430/c
ID AAA91430 standard; DNA; 1599 BP.
XX
AC AAA91430;
XX

D7		12-JUL-2001	(first entry)
XX		Human CHRM5 coding sequence.	
DE			
XX		CHRM5; human; cholinergic receptor muscarinic 5; polymorphic variant;	
KM		genotyping; haplotype; gene therapy; ds.	
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1599	
FT	/**tag=	a	
FT	/product=	"CHRM5"	
FT	replace	(26,G)	
FT	/**tag=	b	
FT	mutation	replace (33,T)	
FT	/**tag=	c	
FT	mutation	replace (345,T)	
FT	/**tag=	d	
FT	*mutation	replace (573,T)	
FT	/**tag=	e	
FT	mutation	replace (1234,A)	
FT	/**tag=	f	
XX			
PN	**MO200128995-A2.		
XX			
PD	26-APR-2001.		
XX			
PF	19-OCT-2000; 2000WO-US29071.		
XX			
PR	21-OCT-1999; 99US-0160647.		
PA	(GENA-) GENAISANCE PHARM INC.		
XX			
PI	Chew A, Chol JY, Nandabalan K, Stephens JC:		
DR	NPI: 2001-300313/31.		
DR	P-PSCDB; AAU97746.		
PT	Isolated polynucleotide encoding the cholinergic receptor, muscarinic 5		
PT	(CHRM5), used to genotype/haplotype the CHRM5 gene, and to identify an		
PT	association between a trait and a polymorphism, comprises novel		
PT	polymorphisms -		
PS	Claim 7; Fig 2; 53pp; English.		
XX			
CC	This sequence encodes the human cholinergic receptor, muscarinic 5		
CC	(CHRM5) protein. The invention relates to a polymorphic variant of this		
CC	CHRM5 sequence. The polymorphic sequence is useful to genotype or		
CC	haplotype the CHRM5 gene, to predict a haplotype pair for the CHRM5 gene,		
CC	and for identifying an association between a trait (such as a clinical		
CC	response to a drug targeting CHRM5). It is also useful in gene therapy in		
CC	patients who lack the CHRM5 isogene or have only one copy of it, and in		
CC	assays to measure the binding affinities of one or more candidate drugs		
CC	targeting CHRM5. The DNA sequence is used in the treatment of disorders		
CC	affected by expression or function of a novel CHRM5 isogene of the		
CC	invention. The protein encoded by the CHRM5 variant is useful to identify		
CC	drugs which target the CHRM5 polymorphic variant protein. Antibodies		
CC	against the protein can be used to neutralise the CHRM5 isoform activity		
CC	expressed in an individual, and is useful in detection of CHRM5 in		
CC	immunocytochemical, immunohistochemical and immunofluorescence. A		
CC	composition containing a genotyping oligonucleotide for detecting a		
CC	polymorphism in the CHRM5 gene is used to detect novel CHRM5		
CC	polymorphisms of the invention.		
SQ	Sequence 1599 BP; 407 A; 455 C; 376 G; 357 T; 0 other:		

Query Match 39.0%; Score 24.6; DB 22; Length 1599;
 Best Local Similarity 70.2%; Pred. No. 30;
 Matches 53; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

17 tgcacatcatctactgcaacaggaagtgttgaaagtcgcagagatg 63

Db	523	TCGCAACCAAGTACTGCGACAGAGAGATTGCTGGGCGCCAGAGAGATG	477
		1 1111 11 111111 11 1111 11 1 1111111111	
RESULT#	9		
AB198000/c			
ID	AB198000	standard; cDNA; 1599 BP.	
AC	AB198000;		
XX			
DT	18-FEB-2002	(first entry)	
XX			
DE	Non-endogenous human GPCR cDNA, SEQ ID NO: 520.		
XX			
XX	Human; G protein-coupled receptor; GPCR; non-endogenous; mutant;		
KW	constitutively activated GPCR; agonist; disease; ss.		
XX			
OS	Homo sapiens.		
XX	Synthetic.		
PN	W0200177172-A2.		
XX			
PD	18-OCT-2001.		
XX			
PF	05-APR-2001; 2001WO-US11098.		
XX			
PR	07-APR-2000; 2000US-195747P.		
XX			
PA	(AREN-) ARENA PHARM INC.		
XX			
PI	Lehmann-Bruhnsma K, Liaw CW, Lin I;		
DR	WPI: 2001-648759/74.		
XX	P-RSDB; ABB56364.		
XX			
PT	Identifying agonists of G protein-coupled receptors (GPCRs) for use in		
PT	disease treatment, comprises contacting candidate compounds with		
XX	versions of GPCRs -		
XX			
XX	Example 2; Page 321-322; 394pp; English.		
XX			
CC	The invention relates to G protein-coupled receptors (GPCRs) for which		
CC	the endogenous ligand has been identified. Non-endogenous		
CC	constitutively activated versions of known GPCRs are used in the		
CC	invention for the direct identification of candidate compounds as		
CC	receptor agonists, inverse agonists or partial agonists. Such		
CC	agonists are useful as therapeutic agents for diseases or disorders		
CC	associated with GPCRs. The present sequence encodes a non-endogenous		
XX	version of a known human GPCR.		
XX			
SO	Sequence 1599 BP; 410 A; 457 C; 375 G; 357 T; 0 other;		
Query Match	39.0%; Score 24.6; DB 23; Length 1599;		
Best Local Similarity	70.2%; Pred. No. 30;		
Matches	33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;		
QY	17 tgcacatcactactgcacaagagagactggtaagtgcacagagatg 63		
	1 1111 11 111111 1111 11 1 1111111111		
Db	523 TCGCAACCAAGTACTGCGACAGAGAGATTGCTGGGCGCCAGAGAGATG 477		
RESULT#	10		
AB198000/c			
ID	AB198000	standard; cDNA; 2253 BP.	
AC	AB198000;		
XX			
DT	17-DEC-2001	(updated)	
XX			
DT	27-SEP-1989	(first entry)	
XX			
DE	Human muscarinic acetylcholine m5 receptor gene.		

XX DE DNA encoding novel human diagnostic protein #28401.
XX XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX MO200175067-A2.
XX PN
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI *Dmanac RT, Liu C, Tang YT;
XX XX
XX WPI; 2001-639362/73.
XX DR P-PSDB: ABG28410.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 1: SEQ ID NO 28401; 103pp; English.
XX XX
XX The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SO Sequence 17745 BP; 4061 A; 5222 C; 4755 G; 3707 T; 0 other;

Query Match 38.7%; Score 24.4; DB 23; Length 17745;
Best Local Similarity 68.0%; Pred. No. 63;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 5 ccaagaagccagtcgaatcatctactgcaacaggaagctgtaagtcg 54
DB 17692 CCAAGAGAGAGTGTGAGCTGAGCTGACGACAGCAGAGACTGTGAGAGC 17643

RESULT 13
ID ABA05165 standard; cDNA; 2604 BP.
XX ABA05165;
XX 04-MAR-2002 (first entry)
XX XX

DE Human gamma glutamyl transpeptidase coding sequence.
XX XX
XX Human; gamma glutamyl transpeptidase; glutamate metabolic defect;
XX Organic acidemia; metabolic disorder; cancer; inflammation; immune disease;
XX embryonic developmental disease; growth dysplasia; immune disease;
XX cytostatic; hemostatic; virucide; immunomodulatory; antiinflammatory;
XX gene therapy; ss.
XX Homo sapiens.
XX OS
XX Key Location/Qualifiers
XX FH 2241..2519
XX FT CDS /tag= a
XX FT /product= "gamma glutamyl transpeptidase"
XX PN
XX MO200187944-A1.
XX PD 22-NOV-2001.
XX XX
XX 28-APR-2001; 2001WO-CN00660.
XX PF
XX PR 29-APR-2000; 2000CN-0115514.
XX PA (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX PI Mao Y, Xie Y;
XX XX
XX WPI; 2002-055687/07.
XX DR P-PSDB: AAM47361.
XX XX
XX Gamma glutamyl transpeptidase and encoding polynucleotide, used in
XX PT diagnosis and treatment of malignant tumors, hemopathy, human
XX PT immunodeficiency virus infection, immunological diseases and
XX PT inflammation -
XX PS Claim 6: Page 31-32; 39pp; Chinese.
XX XX
XX The present invention provides the protein and coding sequences of a
XX CC human gamma glutamyl transpeptidase. The sequences can be used in the
XX CC treatment of cancer, glutamate metabolic defects, organic acidemia,
XX CC metabolic disorders, embryonic developmental diseases, growth
XX CC dysplasia, inflammation and immune diseases. The present sequence is
XX CC the coding sequence of the invention.
XX SO Sequence 2604 BP; 677 A; 550 C; 776 G; 601 T; 0 other;

Query Match 38.4%; Score 24.2; DB 24; Length 2604;
Best Local Similarity 62.3%; Pred. No. 47;
Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 gtcccaagaagccagtcgaatcatctactgcaacaggaagctgtaagtcgcaagaga 61
DB 1627 gtccgaagtagcaatgcaacacactctttagaagatgagtgcaagtggaagggga 1686

QY 62 t 62
DB 1687 t 1687

RESULT 14
ID AAQ40706 standard; DNA; 31122 BP.
XX AAQ40706;
XX 17-AUG-1993 (first entry)
XX DE Bacillus subtilis srfA operon.
XX XX
XX Multienzyme complex; surfactin synthetase; MCSS; ORF; surfactant;
XX KM anti-cholinesterase; fungicidal; antibiotic; anti-coagulant;
XX KM thrombosis; myocardium infarct; pulmonary embolism; ss.

CC (AB857737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 4041 BP; 992 A; 1001 C; 1071 G; 977 T; 0 other;

Query Match 37.5%; Score 23.6; DB 23; Length 4041;
Best Local Similarity 69.6%; Pred. No. 87;
Matches 32; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 15 agtgcacatcatctactgtcaacagagagactgttaagtcagag 60
I
Db 3021 atggccatttgcacaccaccacgacgtgcaatgtgcagtg 3066

Search completed: May 12, 2002, 20:20:39
Job Time: 4926 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 20:13:05 : Search time 1854.92 Seconds
(without alignments)
710.743 Million cell updates/sec

Title: US-09-554-024-2
Perfect score: 63
Sequence: 1 ggttcaagaagcagtgcc.....ctgtaagtgcacagagatcg 63

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_bt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID Description

1	63	100.0	63	6	A95924	A95924 Sequence 2
2	63	100.0	98	6	A95925	A95925 Sequence 3
3	63	100.0	197	6	A95927	A95927 Sequence 5
4	52	82.5	87	6	A95931	A95931 Sequence 9
5	40.4	64.1	42	6	A95930	A95930 Sequence 8
6	33	52.4	33	6	A95923	A95923 Sequence 1
7	27.6	43.8	220257	2	AC005342	AC005342 Sequence 1
8	27.6	43.8	236177	2	AC093660	AC093660 Homo sapi
9	27	42.9	63435	2	AC103764	AC103764 Homo sapi
10	27	42.9	74671	2	AL449106	AL449106 Homo sapi
11	27	42.9	168081	2	AC083955	AC083955 Mus muscu
12	27	42.9	176163	2	AL162255	AL162255 Homo sapi
13	27	42.9	192382	2	AC104697	AC104697 Homo sapi
14	27	42.9	197418	30	AC024403	AC024403 Homo sapi
15	26.8	42.5	1632	10	S75713	S75713 abd-2-aceta
16	26.8	42.5	1659	8	AF275347	AF275347 Cladospor
17	26.8	42.5	2038	10	MUSADPF	M74570 Mouse aldeh
18	26.6	42.2	116181	9	AC010681	AC010681 Homo sapi
19	26.6	41.6	2436	3	AF148508	AF148508 Placodect
20	26.2	41.6	2486	3	AF175578	AF175578 Placodect
21	26.2	41.6	184351	2	AC079783	AC079783 Homo sapi
22	26.2	41.6	190689	2	AC091185	AC091185 Homo sapi
23	26.2	41.6	194543	2	AC027630	AC027630 Homo sapi
24	26	41.3	286550	1	SMES91785	AL591785 Slnorhizo
25	25.8	41.0	106997	2	AC094530	AC094530 Rattus no
26	25.8	41.0	173553	2	AC094561	AC094561 Rattus no
27	25.6	40.6	96489	2	AC021441	AC021441 Homo sapi
28	25.6	40.6	173071	2	AC094125	AC094125 Rattus no
29	25.4	40.3	41631	2	TBCHRIA_09	Continuation (10 o
30	25.2	40.0	2029	5	AF035443	AF035443 Xenopus 1
31	25.2	40.0	22838	1	AF081920	AF081920 Pseudomon
32	25.2	40.0	34158	1	SKL13	AL51267 Streptomy
33	25.2	40.0	84399	2	AC094166	AC094166 Rattus no
34	25.2	40.0	137229	9	AC010877	AC010877 Homo sapi
35	25.2	40.0	139737	9	HS228H13	AL031985 Human DNA
36	25.2	40.0	172079	2	AC098744	AC098744 Papio cyn
37	25.2	40.0	173692	2	AC025894	AC025894 Homo sapi
38	25.2	40.0	175509	2	AC099333	AC099333 Homo sapi
39	25.2	40.0	223725	9	AC005820	AC005820 Homo sapi
40	25.2	40.0	258615	2	AC099306	AC099306 Rattus no
41	25	39.7	110000	2	AC084416	Continuation (4 of
42	25	39.7	122005	9	AC018756	AC018756 Homo sapi
43	25	39.7	124389	2	AP004019	AP004019 Oryza sat
44	25	39.7	139226	9	AC008650	AC008650 Homo sapi
45	25	39.7	150864	9	AC083871	AC083871 Homo sapi

ALIGNMENTS

RESULT 1	A95924	63 bp	DNA	Linear	PAT 26-JAN-2000
LOCUS	A95924	Sequence 2 from Patent WO9924594.			
DEFINITION	A95924				
ACCESSION	A95924				
VERSION	A95924.1	GI:6779828			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
DEPOSE					
FEATURES					
source					

CDS
/organism="unidentified"
/db_xref="taxon:32644"
<1..>63
/note="unnamed protein product"
/codon_start=1

BASE COUNT 19 a 14 c 19 g 11 t
ORIGIN /protein_id="CAB70492.1"
/db_xref="GI:6779829"
/translation="GSKKPVPIIYCNRRGKRCORM"

Query Match 100.0%; Score 63; DB 6; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggttccaagaagccagtcgcaatcatctactgcaacagagagactggttaagtcgcaagag 60
|||||
Db 1 GGTTCGAAGAAGCCAGTCGCAATCATCTACTGCAACAGAGAGACTGTAAGTCGCAAGAG 60

OY 61 atg 63
|||
Db 61 ATG 63

RESULT 2
LOCUS A95925 98 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 3 from Patent WO924594.
ACCESSION A95925
VERSION A95925.1 GI:6779830
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 98)
AUTHORS Derose, R. and Hoffmann, J.

TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 9924594-A 3 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)

FEATURES
source Location/Qualifiers
1..98
/organism="unclassified"
/db_xref="taxon:32644"
CDS
1..66
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70493.1"
/db_xref="GI:6779831"
/translation="GSKKPVPIIYCNRRGKRCORM"

BASE COUNT 25 a 23 c 34 g 16 t
ORIGIN

Query Match 100.0%; Score 63; DB 6; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggttccaagaagccagtcgcaatcatctactgcaacagagagactggttaagtcgcaagag 60
|||||
Db 1 GGTTCGAAGAAGCCAGTCGCAATCATCTACTGCAACAGAGAGACTGTAAGTCGCAAGAG 60

OY 61 atg 63
|||
Db 61 ATG 63

RESULT 3
LOCUS A95927 197 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 5 from Patent WO924594.
ACCESSION A95927
VERSION A95927.1 GI:6779834
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 197)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 9924594-A 5 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)

FEATURES
source Location/Qualifiers
1..197
/organism="unclassified"
/db_xref="taxon:32644"
CDS
12..167
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70495.1"
/db_xref="GI:6779835"
/translation="MCPVLFSQLPSFLVSTLLFLVISHSCRAGSKKPVPIIYCNRR
TKRCORM"

BASE COUNT 32 a 56 c 49 g 60 t
ORIGIN

Query Match 100.0%; Score 63; DB 6; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggttccaagaagccagtcgcaatcatctactgcaacagagagactggttaagtcgcaagag 60
|||||
Db 102 GGTTCGAAGAAGCCAGTCGCAATCATCTACTGCAACAGAGAGACTGTAAGTCGCAAGAG 161

OY 61 atg 63
|||
Db 162 ATG 164

RESULT 4
LOCUS A95931 87 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 9 from Patent WO924594.
ACCESSION A95931
VERSION A95931.1 GI:6779839
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 87)
AUTHORS Derose, R. and Hoffmann, J.

TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 9924594-A 9 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)

FEATURES
source Location/Qualifiers
1..87
/organism="unclassified"
/db_xref="taxon:32644"
BASE COUNT 14 a 31 c 21 g 21 t
ORIGIN

Query Match 82.5%; Score 52; DB 6; Length 87;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 gccagtcgcaatcatctactgcaacagagagactggttaagtcgcaagagatg 63
|||||
Db 87 GCCAGTCGCAATCATCTACTGCAACAGAGAGACTGTAAGTCGCAAGAGATG 36

RESULT 5
LOCUS A95930 42 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 8 from Patent WO924594.
ACCESSION A95930
VERSION A95930.1 GI:6779838

```

KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 42)
AUTHORS     Derose,R. and Hoffmann,J.
TITLE       GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
            TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL     PATENT: WO 9824594-A 8 20-MAY-1999;
            DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
   source
       1..42
       /organism="unidentified"
       /db_xref="taxon:32644"
BASE COUNT  13 a      12 c      10 g      7 t
ORIGIN

Query Match      64.1%; Score 40.4; DB 6; Length 42;
Best Local Similarity 97.6%; Pred. No. 6.8e-05;
Matches 41: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ggtccaaagaccagtcgccaatcattctacgcacacagag 42
    |||||||
DB 1 ggtccaaagaccagtcgccaatcattctacgcacacagag 42

RESULT 6
LOCUS      A95923              33 bp      DNA      linear      PAT 26-JAN-2000
DEFINITION Sequence 1 from Patent WO9924594.
ACCESSION  A95923
VERSION     A95923.1 GI:6779826
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 33)
AUTHORS     Derose,R. and Hoffmann,J.
TITLE       GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
            TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL     PATENT: WO 9924594-A 1 20-MAY-1999;
            DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
   source
       1..33
       /organism="unidentified"
       /db_xref="taxon:32644"
       <1..>33
       /note="unnamed protein product"
       /codon_start=1
       /protein_id="CAB70491.1"
       /db_xref="GI:6779827"
       /translation="ITCNRRRTGKC"
BASE COUNT  10 a      7 c      9 g      7 t
ORIGIN

Query Match      52.4%; Score 33; DB 6; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 33: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 atcatctactgaacagagagactgtatagtc 54
    |||||||
DB 1 atcatctactgaacagagagactgtatagtc 33

RESULT 7
AC005342/c  220257 bp      DNA      linear      PRI 04-DEC-1998
LOCUS      AC005342
DEFINITION Homo sapiens 12p13.3 PAC RPI15-1096D14 (Roswell Park Cancer
            Institute Human PAC Library) complete sequence.
ACCESSION  AC005342
VERSION     AC005342.1 GI:3810573

```

```

KEYWORDS
SOURCE      HMC.
ORGANISM    human.
REFERENCE   1 (bases 1 to 220257)
AUTHORS     Muzny,D., Aronson,A.D., Bouck,J., Bunac,C., Chang,J., Chen,Z.,
            Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
            Gorrell,L.L., Hernandez,J., Jackson,L., Kondogewski,N., Lau,S.,
            Leal,B., Lee,E., Lichtarge,O., Liu,M., Logan,O., Lu,J.,
            Marandel,I., Martinez,C., Merscher,S., Montgomery,K., Oswal,G.,
            Pampell,L.R., Parish,B.J., Perez,L., Rashid,N.D., Renault,B.,
            Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,Q., Williamson,A.,
            Worley,K.C., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and
            Gibbs,R.A.
TITLE       Direct Submission
JOURNAL     unpublished
REFERENCE   2 (bases 1 to 220257)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (31-JUL-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 220257)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (30-OCT-1998) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 220257)
Worley,K.C.
Direct Submission
Submitted (04-DEC-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 30, 1998 this sequence version replaced g1:3723954.
On Oct 30, 1998 this sequence replaced g1:3723954.
INFORMATION: http://gc.bcm.tmc.edu:8088/home.html or email
gc-help@bcm.tmc.edu

```

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-14) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----

Contig Length: 220257
 Phrap values in estimate: 58222
 Average error rate (RCM-phrap estimate): 9.8888e-07
 Fraction of Phrap values less than 40 : 0.00173474
 Number of consensus changing edits: 10
 Number of N's in consensus : 0

Position	Original+Context	Edited+Context
33512	attctttat(n)tatattatta	attgtttat(t)tatattatta
54816	acacaaaga(n)aataccaga	acacaaaga(a)aataccaga
56795	cgtgtgttac(n)caaggtcga	cgtgtgttac(c)caaggtcga
120971	aaaaaaa(c)caacaactc	aaaaaaa(a)caacaactc
152487	ttcaaaaga(n)actltagaag	ttcaaaaga(a)actltagaag
176875	ggggtcgaag(n)ccggtcgag	ggggtcgaag(g)ccggtcgag
176918	cgagagggg(n)gttcgcgc	cgagagggg(t)gttcgcgc
177533	ccgcctcctc(n)ttcgcgc	ccgcctcctc(c)ttcgcgc
178001	ccgcctcctc(n)ggggtcgag	ccgcctcctc(c)ggggtcgag
195474	ccctccctcc(n)ttcctccctc	ccctccctcc(c)ttcctccctc

----- Low Quality Bases -----

Quality Position Surrounding Sequence

----- Distribution of Quality < 40 Bases -----

#	bases	5	10	15	20	25	30	35	40
1001	1001								
901	901								
801	801								
701	701								
601	601								
501	501								
401	401								
301	301								
201	201								
101	101								
01	01								

Phrap Value Range

Version: 1.0 txio.

FEATURES Location/Qualifiers

source 1..220257 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="12p13.3"
 /clone="RPC15-1096D14"
 misc_feature 1..78 /note="Bases 1 to 16417 overlaps the adjacent clone RPC11-21K20 only 78 bases of overlap included in AC005343."
 /function="Overlap with adjacent clone AC005343."
 misc_difference 2593 /note="Insertion of A."
 /clone="Insertion with respect to AC005343."
 repeat_region complement(3901..4051)
 /rpt_family="MER58"
 misc_difference 6497 /note="A in AC005343."
 /clone="Substitution with respect to AC005343."
 repeat_region 7931..7979 /rpt_family="L2"
 8136..8227 /rpt_family="L2"
 repeat_region /rpt_family="L2"

misc_difference 8487 /note="C in AC005343."
 /clone="Substitution with respect to AC005343."
 misc_difference 10553 /note="C in AC005343."
 /clone="Substitution with respect to AC005343."
 misc_difference 10772 /note="C in AC005343."
 /clone="Substitution with respect to AC005343."
 misc_difference 12486 /note="G in AC005343."
 /clone="Substitution with respect to AC005343."
 repeat_region 13057..13629 /rpt_family="L1M4"
 misc_difference 13194 /note="G in AC005343."
 /clone="Substitution with respect to AC005343."
 misc_difference 13419 /note="C in AC005343."
 /clone="Substitution with respect to AC005343."
 repeat_region complement(13813..13922)
 /rpt_family="FLAM_A"
 repeat_region 14162..17576 /rpt_family="L1M4"
 misc_difference 14448 /note="C in AC005343."
 /clone="Substitution with respect to AC005343."
 misc_difference 15248 /note="G in AC005343."
 /clone="Substitution with respect to AC005343."
 misc_difference 16418 /note="G in AC005343."
 /clone="Insertion with respect to AC005343."
 misc_difference 16421..16427 /note="GTTCTTC in AC005343."
 /clone="Insertion with respect to AC005343."
 misc_difference 16431..16435 /note="GTCCA in AC005343."
 /clone="Insertion with respect to AC005343."
 misc_difference 16464..16467 /note="CGCG in AC005343."
 /clone="Substitution with respect to AC005343."
 repeat_region 17577..17853 /rpt_family="A1usg1"
 repeat_region 17855..19931 /rpt_family="L1M1"
 repeat_region complement(19962..20254)
 /rpt_family="A1uud0"
 repeat_region 20261..21347

Query Match 43.8%; Score 27.6; DB 9; Length 220257;
 Best Local Similarity 67.2%; Pred. No. 12;
 Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 2 gttccaagaagccatgccaatcatctatgcaacagagagactgttaagccagag 59
 DB 98878 GCTCTTACTTGCAGTGCCTAACAAAGTACTCAATCAGCAAGATGTAATGCTAGAG 98821

RESULT 8
 AC093660 236177 bp DNA linear HTG 24-JAN-2002
 LOCUS Homo sapiens chromosome 4 clone RP11-470D11, WORKING DRAFT
 DEFINITION SEQUENCE, 23 unordered pieces.
 ACCESSION AC093660
 VERSION AC093660.2 GI:18308900
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 236177)
 AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 236177)
AUTHORS Waterston, R.H.
TITLE Direct Submission
Submitted (07-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jan 24, 2002 this sequence version replaced gi:15487495.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H.NH0470D11
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 228670 bases at least Q40
Consensus quality: 232856 bases at least Q30
Insert size: 171000; agarose-fp
Insert size: 234376; sum-of-contigs
Quality coverage: 11.97 in Q20 bases; 8.83 in Q20 bases; sum-of-contigs
Quality coverage: 8.83 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1169: contig of 1169 bp in length
1170 1269: gap of unknown length
1270 1395: contig of 1126 bp in length
2396 2495: gap of unknown length
2496 3610: contig of 1115 bp in length
3611 3710: gap of unknown length
3711 4819: contig of 1109 bp in length
4820 4919: gap of unknown length
4920 6071: contig of 1152 bp in length
6072 6171: gap of unknown length
6172 7831: contig of 1660 bp in length
7832 7931: gap of unknown length
7932 9100: contig of 1169 bp in length
9101 9200: gap of unknown length
9201 10495: contig of 1295 bp in length
10496 10595: gap of unknown length
10596 11673: contig of 1078 bp in length
11674 11774: gap of unknown length
11775 13026: contig of 1253 bp in length
13027 13126: gap of unknown length
13127 14937: contig of 1811 bp in length
14938 15037: gap of unknown length
15038 16737: contig of 1700 bp in length
16738 16837: gap of unknown length
16838 18422: contig of 1585 bp in length
18423 18522: gap of unknown length
18523 19834: contig of 1312 bp in length
19835 19934: gap of unknown length
19935 22386: contig of 2652 bp in length
22387 22686: gap of unknown length
22687 25068: contig of 2382 bp in length
25069 25168: gap of unknown length
25169 27681: contig of 2513 bp in length

FEATURES
source * 27682 27781: gap of unknown length
* 27782 29591: contig of 1810 bp in length
* 29592 29691: gap of unknown length
* 29692 38111: contig of 8420 bp in length
* 38112 38211: gap of unknown length
* 38212 47759: contig of 9548 bp in length
* 47760 47859: gap of unknown length
* 47860 61458: contig of 13599 bp in length
* 61459 61558: gap of unknown length
* 61559 235391: contig of 173833 bp in length
* 235392 235491: gap of unknown length
* 235492 236177: contig of 686 bp in length.
Location/Qualifiers
1. .236177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-470D11"
1. .1169
/note="assembly_name:Contig44"
1270. .2395
/note="assembly_name:Contig50"
2496. .3610
/note="assembly_name:Contig56"
3711. .4819
/note="assembly_name:Contig62"
4920. .6071
/note="assembly_name:Contig63"
6172. .7831
/note="assembly_name:Contig65"
7932. .9100
/note="assembly_name:Contig68"
9201. .10495
/note="assembly_name:Contig70"
10596. .11673
/note="assembly_name:Contig74"
11774. .13026
/note="assembly_name:Contig75"
13127. .14937
/note="assembly_name:Contig76"
15038. .16737
/note="assembly_name:Contig77"
16838. .18422
/note="assembly_name:Contig78"
18523. .19834
/note="assembly_name:Contig79"
19935. .22586
/note="assembly_name:Contig80"
22687. .25068
/note="assembly_name:Contig81"
25169. .27681
/note="assembly_name:Contig82"
27782. .29591
/note="assembly_name:Contig84"
29692. .38111
/note="assembly_name:Contig85"
38212. .47759
/note="assembly_name:Contig86"
47860. .61458
/note="assembly_name:Contig87"
61559. .235391
/note="assembly_name:Contig88
clone_end:T7
vector_side:left"
235492. .236177
/note="assembly_name:Contig4"
BASE COUNT 64649 a 51668 c 51720 g 65858 t 2282 others
ORIGIN
Query Match 43.8%; Score 27.6; DB 2; Length 236177;
Best Local Similarity 67.2%; Pred. No. 13;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

0Y 2 gttccagaagcagatgacatcatctactgcacagagagactgtaagtcacagag 59
DB 6441 GCTCTTAGTTGCGACGCTCAACACAGTACATCAAGAGAAATGATATGTGTAGAG 6498
RESULT 9
AC103764/c
LOCUS
DEFINITION Homo sapiens chromosome 8 clone RP11-1057B8 map 8, LOW-PASS
AC103764 63435 bp DNA linear HTG 29-NOV-2001
SEQUENCE SAMPLING:
AC103764
AC103764.1 GI:17149649
HTG; HTGS_PHASE0.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 63435)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 8, clone RP11-1057B8
Unpublished
2 (bases 1 to 63435)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fairo, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heald, A., Horton, L., Hulme, W., Illey, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,
Lamazan, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McKean, P., McKernan, K., McPheters, R., Meldrum, J.,
Meneu, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zalouni, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 121724
Center clone name: 1057_B_8

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 656 755: gap of 100 bp
* 656 755: contig of 655 bp in length

756 1426: contig of 671 bp in length
1427 1526: gap of 100 bp
1527 2169: contig of 643 bp in length
2170 2269: gap of 100 bp
2270 2943: contig of 674 bp in length
2944 3043: gap of 100 bp
3044 3713: contig of 670 bp in length
3714 3813: gap of 100 bp
3814 4454: contig of 641 bp in length
4455 4554: gap of 100 bp
4555 5219: contig of 665 bp in length
5220 5319: gap of 100 bp
5320 5979: contig of 660 bp in length
5980 6079: gap of 100 bp
6080 6731: contig of 652 bp in length
6732 6831: gap of 100 bp
6832 7511: contig of 680 bp in length
7512 7611: gap of 100 bp
7612 8289: contig of 678 bp in length
8290 8389: gap of 100 bp
8390 9056: contig of 667 bp in length
9057 9156: gap of 100 bp
9157 9820: contig of 664 bp in length
9821 9920: gap of 100 bp
9921 10565: contig of 665 bp in length
10566 10685: gap of 100 bp
10686 11354: contig of 669 bp in length
11355 11454: gap of 100 bp
11455 12115: contig of 661 bp in length
12116 12215: gap of 100 bp
12216 12895: contig of 680 bp in length
12896 12995: gap of 100 bp
12996 13648: contig of 653 bp in length
13649 13748: gap of 100 bp
13749 14419: contig of 671 bp in length
14420 14519: gap of 100 bp
14520 15201: contig of 682 bp in length
15202 15301: gap of 100 bp
15302 15927: contig of 626 bp in length
15928 16027: gap of 100 bp
16028 16713: contig of 686 bp in length
16714 16813: gap of 100 bp
16814 17478: contig of 665 bp in length
17479 17578: gap of 100 bp
17579 18233: contig of 655 bp in length
18234 18333: gap of 100 bp
18334 19006: contig of 673 bp in length
19007 19106: gap of 100 bp
19107 19776: contig of 670 bp in length
19777 19876: gap of 100 bp
19877 20551: contig of 675 bp in length
20552 20651: gap of 100 bp
20652 21327: contig of 676 bp in length
21328 21427: gap of 100 bp
21428 22090: contig of 663 bp in length
22091 22190: gap of 100 bp
22191 22881: contig of 691 bp in length
22882 22981: gap of 100 bp
22982 23627: contig of 646 bp in length
23628 23727: gap of 100 bp
23728 24397: contig of 670 bp in length
24398 24497: gap of 100 bp
24498 25160: contig of 663 bp in length
25161 25260: gap of 100 bp
25261 25941: contig of 681 bp in length
25942 26041: gap of 100 bp
26042 26692: contig of 651 bp in length
26693 26792: gap of 100 bp
26793 27459: contig of 667 bp in length
27460 27559: gap of 100 bp
27560 28227: contig of 668 bp in length
28228 28327: gap of 100 bp
28328 28976: contig of 649 bp in length

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 168081)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL Mus musculus chromosome, clone RP23-84P17
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 168081)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Barnum, N., Bastien, V., Beda, F., Boguslavsky, L.,
Boutgaltier, B., Brown, A., Burkett, G., Campiano, A., Castle, A.,
Choepeil, Y., Colangelo, M., Collins, S., Collinmore, A., Cooke, P.,
D'Arlemano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P.,
Fitzhugh, W., Gage, D., Galagan, J., Gatoyna, S., Gimble, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lacroque, K.,
Lamazars, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrim, J., Menus, L., Mihova, T., Mlenga, V.,
Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K.,
Pierre, N., Pisanic, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
Rogor, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Sudamanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Travers, M., Trijillo, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (08-OCT-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 22, 2000 this sequence version replaced gi:10717222.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: l11090
Center clone name: 84_P_17
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 147448 bases at least Q40
Consensus quality: 158296 bases at least Q30
Consensus quality: 162516 bases at least Q20
Insert size: 198000; agarose-1p
Insert size: 164581; sum-of-ctnigs
Quality coverage: 2.8 in Q20 bases; agarose-1p
Quality coverage: 3.4 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 8071: ctnig of 8071 bp in length
* 8072 8171: gap of 100 bp
* 8172 8798: ctnig of 627 bp in length
* 8799 8898: gap of 100 bp
* 8899 9551: ctnig of 653 bp in length
* 9552 9651: gap of 100 bp
* 9652 10408: ctnig of 757 bp in length
* 10409 10508: gap of 100 bp
* 10509 11835: ctnig of 1327 bp in length

11836 11935: gap of 100 bp
11936 13013: ctnig of 1078 bp in length
13014 13113: gap of 100 bp
13114 14364: ctnig of 1271 bp in length
14365 14484: gap of 100 bp
14485 15189: ctnig of 705 bp in length
15190 15289: gap of 100 bp
15290 16626: ctnig of 1337 bp in length
16627 16726: gap of 100 bp
16727 17872: ctnig of 1146 bp in length
17873 17972: ctnig of 100 bp
17973 19113: ctnig of 1141 bp in length
19114 19213: gap of 100 bp
19214 20332: ctnig of 1119 bp in length
20333 20432: gap of 100 bp
20433 22596: ctnig of 2164 bp in length
22597 22696: gap of 100 bp
22697 25492: ctnig of 2796 bp in length
25493 25592: gap of 100 bp
25593 28554: ctnig of 2962 bp in length
28555 28654: gap of 100 bp
28655 31898: ctnig of 3244 bp in length
31899 31998: gap of 100 bp
31999 35948: ctnig of 3850 bp in length
35949 35948: gap of 100 bp
35949 39030: ctnig of 3082 bp in length
39031 39130: gap of 100 bp
39131 42611: ctnig of 3461 bp in length
42612 42711: gap of 100 bp
42712 45681: ctnig of 2970 bp in length
45682 45781: gap of 100 bp
45782 48839: ctnig of 3058 bp in length
48840 48939: gap of 100 bp
48940 53014: ctnig of 4075 bp in length
53015 53114: gap of 100 bp
53115 56891: ctnig of 3777 bp in length
56892 56991: gap of 100 bp
56992 62353: ctnig of 5362 bp in length
62354 62453: gap of 100 bp
62454 66132: ctnig of 3679 bp in length
66133 66232: gap of 100 bp
66233 70771: ctnig of 4539 bp in length
70772 70871: gap of 100 bp
70872 76244: ctnig of 5373 bp in length
76245 76344: gap of 100 bp
76345 80719: ctnig of 4375 bp in length
80720 80819: gap of 100 bp
80820 85869: ctnig of 5050 bp in length
85870 85969: gap of 100 bp
85970 92336: ctnig of 6427 bp in length
92337 92436: gap of 100 bp
92437 99048: ctnig of 6552 bp in length
99049 99148: gap of 100 bp
99149 109053: ctnig of 9905 bp in length
109054 109153: gap of 100 bp
109154 120325: ctnig of 11172 bp in length
120326 120425: gap of 100 bp
120426 132250: ctnig of 11825 bp in length
132251 132350: gap of 100 bp
132351 145705: ctnig of 13355 bp in length
145706 145805: gap of 100 bp
145806 168081: ctnig of 22276 bp in length.
Location/Qualifiers
1. 168081
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-84P17"
/clone_lib="RP23-84P17"
1. 8071
/note="assembly-fragment"
clone_end:Sp6
vector_side:left"
8172. 8798
misc_feature
misc_feature
misc_feature

```

misc_feature      /note="assembly-fragment"
8899..9551
misc_feature      /note="assembly-fragment"
9652..10408
misc_feature      /note="assembly-fragment"
10509..11835
misc_feature      /note="assembly-fragment"
11936..13013
misc_feature      /note="assembly-fragment"
13114..14384
misc_feature      /note="assembly-fragment"
14485..15189
misc_feature      /note="assembly-fragment"
15290..16626
misc_feature      /note="assembly-fragment"
16727..17872
misc_feature      /note="assembly-fragment"
17973..19113
misc_feature      /note="assembly-fragment"
19214..20332
misc_feature      /note="assembly-fragment"
20433..22596
misc_feature      /note="assembly-fragment"
22697..25492
misc_feature      /note="assembly-fragment"
25593..28554
misc_feature      /note="assembly-fragment"
28655..31898
misc_feature      /note="assembly-fragment"
31999..35848
misc_feature      /note="assembly-fragment"
35949..39030
misc_feature      /note="assembly-fragment"
39131..42611
misc_feature      /note="assembly-fragment"
42712..45681
misc_feature      /note="assembly-fragment"
45782..48839
misc_feature      /note="assembly-fragment"
48940..53014
misc_feature      /note="assembly-fragment"
53115..56891
misc_feature      /note="assembly-fragment"
56992..62353
misc_feature      /note="assembly-fragment"
62454..66132
misc_feature      /note="assembly-fragment"
66233..70771

Query Match      42.9%: Score 27; DB 2; Length 168081;
Best Local Similarity 70.6%; Pred. No. 21;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY      3      ttccaaagagccagtcgaatcatctactgcacaagagagactgtaagt 53
Db 164047 TTGATAGAAAGTCGGCCATTCATCTGTAAGAGGAGAAAGGAGGAGTG 163997

RESULT 12
AL162255      176163 bp      DNA      linear      HTG 12-FEB-2002
LOCUS      Homo sapiens chromosome 1 clone RP11-18E13 map q24.3-25.3, ***
DEFINITION      SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION      AL162255
VERSION      AL162255.16 GI:16655963
KEYWORDS      HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 176163)
AUTHORS      Harrison, E.
TITLE      Direct Submission

```

```

JOURNAL      Submitted (11-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 13, 2002 this sequence version replaced g1:18250452.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA18E13
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 175707 bases at least Q40
Consensus quality: 175950 bases at least Q40
Consensus quality: 176098 bases at least Q30
Insert size: 176163; sum-of-contigs
Insert size: 172792; agarose-fp
Quality coverage: 7.21x in Q20 bases; sum-of-contigs quality
coverage: 7.90x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..176163
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q24.3-25.3"
/clone="RP11-18E13"
/clone_1fb="RPC1-11.1"
1..176163
/note="assembly-fragment:02749
clone_end:=f7
vector_side:left"

BASE COUNT      51776 a 32454 c 34988 g 56945 t
ORIGIN

Query Match      42.9%: Score 27; DB 2; Length 176163;
Best Local Similarity 70.6%; Pred. No. 21;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY      2      gtccaaagagccagtcgaatcatctactgcacaagagagactgtaagt 52
Db 29957 GTTCAAGTACACACTTCATCTGTTGACAGGATCATTTGTAAAGT 30007

RESULT 13
AC104697      192382 bp      DNA      linear      HTG 19-DEC-2001
LOCUS      Homo sapiens chromosome 2 clone RP11-454B3, WORKING DRAFT SEQUENCE,
2 unordered pieces.
ACCESSION      AC104697 AC024403
VERSION      AC104697.1 GI:17933873
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 192382)
AUTHORS      Waterston, R.H.
TITLE      The sequence of Homo sapiens clone
unpublished
JOURNAL      2 (bases 1 to 192382)
AUTHORS      Waterston, R.H.
REFERENCE      Direct Submission
TITLE      Submitted (19-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```

DE Homo sapiens chromosome 2 clone RP11-454B3 map 2, WORKING DRAFT SEQUENCE.
DE
DE 44 unordered pieces.
XX HTGS_DRAFT; HTGS_PHASE1.
XX
OC Homo sapiens (human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
XX
XX [1]
RP 1-197418
RA Birren B., Linton L., Nusbaum C., Lander E.,
RL "Homo sapiens chromosome 2, clone RP11-454B3";
unpublished.
XX
XX [2]
RP 1-197418
RA Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
RA Anderson S., Baldwin J., Barna N., Bede F., Boguslavsky L., Boukhalter B.,
RA Brown A., Burkett G., Campopiano A., Castile A., Choepel Y., Colangelo M.,
RA Collins S., Collamore A., Cooke P., Dearlano K., Dewar K., Diaz J.S.,
RA Dodge S., Donlo M., Doyle M., Fenesor J., Ferreira P., Fitzhugh W.,
RA Forrest C., Gage D., Galagan J., Gardyna S., Ginde S., Goyette M.,
RA Graham L., Grand-Pierre N., Grant G., Hagos B., Healdorf A., Horton L.,
RA Howland J.C., Iliev I., Johnson R., Jones C., Kann L., Karatas A.,
RA Klein J., Landers T., Largocque K., Lehoczy J., Levine R., Lieu C.,
RA Liu G., Locke K., Macdonald P., Marcus N., McCarthy M., McEwan P.,
RA McCracken K., McKernan K., McPheters R., Meldrum J., Menais T., Mihova T.,
RA Miranda C., Mlenka V., Morrow J., Naylor J., Norman C.H., O'Connor T.,
RA O'Donnell P., O'Neil D., Oliver T.M., Peterson K., Pierre N., Pisani C.,
RA Pollara V., Raymond C., Riley R., Rogov P., Rothman D., Roy A., Santos R.,
RA Schrauer S., Severy P., Spencer B., Stange-Thoman N., Stojanovic N.,
RA Subramanian A., Talamas J., Testlay S., Theodore J., Tirrell A.,
RA Travers M., Triaillo J., Vassilay H., Viel R., Vo A., Wilson B., Wu X.,
RA Wyman D., Ye W.J., Young G., Zainoun J., Zimmer A., Zody M.;
!
RT Submitted (28-FEB-2000) to the EMBL/GenBank/DDBJ databases.
RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL Cambridge, MA 02141, USA
XX
XX On Mar 12, 2000 this sequence version replaced gi:7108199.
CC All repeats were identified using RepeatMasker:
CC Smit, A.F.A. & Green, P. (1996-1997)
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC ----- Genome Center
CC
CC Center: Whitehead Institute/ MIT Center for Genome Research
CC
CC Center code: WIBR
CC Web site: http://www.seq.wi.mit.edu
CC Contact: sequence_submissions@genome.wi.mit.edu
CC ----- Project Information
CC
CC Center project name: L6834
CC
CC Center clone name: 454_B3
CC
CC ----- Summary Statistics
CC
CC Sequencing vector: M13; M77815; 100% of reads
CC Chemistry: dye-terminator Big Dye; 100% of reads
CC Assembly program: Phrap; version 0.960731
CC Consensus quality: 157857 bases at least Q40
CC Consensus quality: 176423 bases at least Q30
CC Consensus quality: 186150 bases at least Q20
CC Insert size: 190000; agarose-fp
CC
CC Quality coverage: 3.2 in Q20 bases; agarose-fp
CC Quality coverage: 3.2 in Q20 bases; sum-of-contigs
CC -----
CC * NOTE: This is a 'working draft' sequence. It currently
CC consists of 44 contigs. The true order of the pieces
CC is not known and their order in this sequence record is
CC arbitrary. Gaps between the contigs are represented as
CC runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence
CC as soon as it is available and the accession number will
CC be preserved.
CC
CC 1 1100: contig of 1100 bp in length

CC	*	1101	1200:	gap of	100 bp	in	length
CC	*	1201	2438:	contig of	1238 bp	in	length
CC	*	2439	2558:	gap of	100 bp	in	length
CC	*	2539	3781:	contig of	1243 bp	in	length
CC	*	3782	3881:	gap of	100 bp	in	length
CC	*	5045	5045:	contig of	1164 bp	in	length
CC	*	5145	5145:	gap of	100 bp	in	length
CC	*	5146	6642:	contig of	1497 bp	in	length
CC	*	6643	6742:	gap of	100 bp	in	length
CC	*	6743	8164:	contig of	1422 bp	in	length
CC	*	8165	8264:	gap of	100 bp	in	length
CC	*	8265	9726:	contig of	1462 bp	in	length
CC	*	9727	9826:	gap of	100 bp	in	length
CC	*	9827	11739:	contig of	11913 bp	in	length
CC	*	11740	11839:	gap of	100 bp	in	length
CC	*	11840	13330:	contig of	1431 bp	in	length
CC	*	13331	13430:	gap of	100 bp	in	length
CC	*	13431	15649:	contig of	2219 bp	in	length
CC	*	15650	15749:	gap of	100 bp	in	length
CC	*	15750	18374:	contig of	2635 bp	in	length
CC	*	18375	18474:	gap of	100 bp	in	length
CC	*	18475	20648:	contig of	2174 bp	in	length
CC	*	20649	20748:	gap of	100 bp	in	length
CC	*	20749	22015:	contig of	1267 bp	in	length
CC	*	22016	22115:	gap of	100 bp	in	length
CC	*	22116	24565:	contig of	2741 bp	in	length
CC	*	24857	24956:	gap of	100 bp	in	length
CC	*	24957	28230:	contig of	3274 bp	in	length
CC	*	28231	28330:	gap of	100 bp	in	length
CC	*	28331	30774:	contig of	2444 bp	in	length
CC	*	30775	30874:	gap of	100 bp	in	length
CC	*	30875	34408:	contig of	3534 bp	in	length
CC	*	34409	34508:	gap of	100 bp	in	length
CC	*	34509	36298:	contig of	1790 bp	in	length
CC	*	36299	36398:	gap of	100 bp	in	length
CC	*	36399	38662:	contig of	2264 bp	in	length
CC	*	38663	38762:	gap of	100 bp	in	length
CC	*	38763	42095:	contig of	3333 bp	in	length
CC	*	42096	42195:	gap of	100 bp	in	length
CC	*	42196	45055:	contig of	2860 bp	in	length
CC	*	45056	45155:	gap of	100 bp	in	length
CC	*	45156	47782:	contig of	2677 bp	in	length
CC	*	47783	47882:	gap of	100 bp	in	length
CC	*	47883	52304:	contig of	4422 bp	in	length
CC	*	52305	52404:	gap of	100 bp	in	length
CC	*	52405	55828:	contig of	3424 bp	in	length
CC	*	55829	55928:	gap of	100 bp	in	length
CC	*	55929	59658:	contig of	3730 bp	in	length
CC	*	59659	59758:	gap of	100 bp	in	length
CC	*	59759	63879:	contig of	4221 bp	in	length
CC	*	63880	64079:	gap of	100 bp	in	length
CC	*	64080	67509:	contig of	3430 bp	in	length
CC	*	67510	67609:	gap of	100 bp	in	length
CC	*	67610	72512:	contig of	4993 bp	in	length
CC	*	72513	72612:	gap of	100 bp	in	length
CC	*	72613	77741:	contig of	5129 bp	in	length
CC	*	77742	77841:	gap of	100 bp	in	length
CC	*	77842	83349:	contig of	5508 bp	in	length
CC	*	83350	83449:	gap of	100 bp	in	length
CC	*	83450	88863:	contig of	5514 bp	in	length
CC	*	88964	89063:	gap of	100 bp	in	length
CC	*	89064	93791:	contig of	4728 bp	in	length
CC	*	93792	93891:	gap of	100 bp	in	length
CC	*	93892	98121:	contig of	4220 bp	in	length
CC	*	98122	98221:	gap of	100 bp	in	length
CC	*	98222	103190:	contig of	4969 bp	in	length
CC	*	103191	103290:	gap of	100 bp	in	length
CC	*	103191	109512:	contig of	6222 bp	in	length
CC	*	109513	109612:	gap of	100 bp	in	length
CC	*	109613	115504:	contig of	3882 bp	in	length
CC	*	115505	115604:	gap of	100 bp	in	length
CC	*	115605	121119:	contig of	5515 bp	in	length
CC	*	121120	121219:	gap of	100 bp	in	length

CC	*	121320	1325578:	contig of 4359 bp	in length
CC	*	125579	125678:	gap of	100 bp
CC	*	125679	131585:	contig of 3507 bp	in length
CC	*	131586	131685:	gap of	100 bp
CC	*	131686	138039:	contig of 6354 bp	in length
CC	*	138040	138139:	gap of	100 bp
CC	*	138140	148314:	contig of 10075 bp	in length
CC	*	148215	148314:	gap of	100 bp
CC	*	148315	158365:	contig of 10051 bp	in length
CC	*	158366	158465:	gap of	100 bp
CC	*	158466	169691:	contig of 11226 bp	in length
CC	*	169692	169791:	gap of	100 bp
CC	*	169792	197418:	contig of 27627 bp	in length.
XX					
Key		Location/Qualifiers			
FT		source	1. 197418		
FT			/chromosome="2"		
FT			/db_xref="taxon:9606"		
FT			/organism="Homo sapiens"		
FT			/map="2"		
FT			/clone="RP11-454B3"		
FT		misc_feature	1. 1100		
FT			/clone_lib="RCR1-11 Human Male BAC"		
FT		misc_feature	/note="assembly_fragment"		
FT			1201. 2438		
FT		misc_feature	/note="assembly_fragment"		
FT			2539. 3781		
FT		misc_feature	/note="assembly_fragment"		
FT			3882. 5045		
FT		misc_feature	/note="assembly_fragment"		
FT			5146. 6642		
FT		misc_feature	/note="assembly_fragment"		
FT			6743. 8164		
FT		misc_feature	/note="assembly_fragment"		
FT			8265. 9726		
FT		misc_feature	/note="assembly_fragment"		
FT			9827. 11759		
FT		misc_feature	/note="assembly_fragment"		
FT			11840. 13330		
FT		misc_feature	/note="assembly_fragment"		
FT			13431. 15649		
FT		misc_feature	/note="assembly_fragment"		
FT			15750. 18374		
FT		misc_feature	/note="assembly_fragment"		
Query Match			42.9%:	Score 27;	DB 30; Length 197418;
Best Local Similarity			76.7%:	Pred. No. 21;	
Matches 33;		Conservative	0;	Mismatches	10; Indels 0; Gaps 0;
QY	3	ttccaagaagccagtcgcatcatcattactgacacagagact	45		
Db	161203	TTCTGAGATGCAGTGCATATATCTACTGTTATCTGAGAGACT	161161		
RESULT 15					
LOCUS	S75713		1632 bp	mRNA	linear
DEFINITION	ahd-2-acetaldehyde dehydrogenase [mice, BALB/c, livers, mRNA				
	partial, 1632 nt].				
ACCESSION	S75713				
VERSION	S75713.1				
KEYWORDS	GI:913809				
SOURCE	house mouse BALB/c livers.				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 1632)				
AUTHORS	Bond,S.L. and Singh,S.M.				
TITLE	DNA sequence analysis of the cytosolic acetaldehyde dehydrogenase				
	gene (Ahd-2) in mouse strains with variable ethanol preferences				
JOURNAL	Biochem. Med. Metab. Biol. 52 (2), 155-159 (1994)				
MEDLINE	95085815				

REMARK

GenBank staff at the National Library of Medicine created this entry (NCBI gibbsg 161480) from the original journal article. This sequence comes from Fig. 1A.

FEATURES

Source

Location/Qualifiers
1. .1632

```
/organism="Mus musculus"
/strain="BATB/c"
```

```
/db xref="taxon:10090"
```

```
/usr_xr= /usr_xr:10050
/tissue_type="liver"
```

1. .1632

3

```
/gene="ahd-2"
```

```
/gene="ahd-2"
```

/note="This sequence comes from Fig. 1A; conceptual translation presented here differs from translation in publication"; ALDH"

```
/codon_start=1
/product="acet
```

```
/product="acetaldehyde dehydrogenase  
/protein_id="AAB32754.2"
```

```
/procinfo_id= MADS2/04:2
/db_xref="GI:9755362"
```

/translation="MSSPAQP

/t/ia/s/i/lo/ "USSPOAPVAPDILTOHTKIFINEMHNNSVSGKRPPLDRA
 SEEVICUOEBGRKAVDKVAKARAOFOIGSPMRTDASRGRTLALMEPDRILL
 LATMALINGOKFANAYSLDGICIKALCYACMAKIHOOITPSDITFETPREET
 GYCCHGIPNNFPLMFTIKTIGPALSNGVNYVYVPAQDPIHLIASLIEAGPBPSC
 VNIYVEIGTSPAGALISHMDVATATGSOVLKREAGKSULKTYLLELGSPOC
 IYFADADILIEFAHGFVHOQCQVAVASRFEVSZYDEFKRFVPEKPAKVLNKA
 PLTPBINOQDIDKEHGLTILDIIESGKEGALIECGGSMWNGKFFVOPRENSVYD
 EKRILKEBTEFGVOQIMKRSYDVITIRANTTYVGLAAGFTLDDOLATVSSALQOG
 VYVWACVYMLASLQCPGGRKMSNGEHLGHGTYEATLELTVAMKISOXNS"

BASE COUNT	460 a	341 c	426 g	405 t
ORIGIN				

[illegible]

42.58; Score 26.8; DB 10; Length 1632;

Best Local Similarity	68.5%;	Pred. NO. 16;							
Matches	37:	Conservative	0:	Mismatches	17:	Indels	0:	Gaps	0:

MAKING	OF	CONSTRUCTIVE	OF	MINORITIES	OF	INDUST	OF	WAYS	OF
MAKING	OF	CONSTRUCTIVE	OF	MINORITIES	OF	INDUST	OF	WAYS	OF

QY 10 aagccagtgcacatctactgcacacagagagactggtagtgcacagagatg 63

Db 551 AGGGCAGGGCTATCTTCCAATGAACATGAGCATTGGAATAATCCAGGGATG 498

```
Search completed: May 12, 2002, 20:13:53
Job time: 7215 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 19:41:35 ; Search time 2487.21 Seconds
(without alignments)
1069.029 Million cell updates/sec

Title: US-09-554-024-5

Perfect score: 197
Sequence: 1 gcgtcgcgcgcgtcgttc.....gcgaacgtcgcgcgcgttc 197

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estlba:*
2: em_estlhum:*
3: em_estlin:*
4: em_estlmu:*
5: em_estlro:*
6: em_estlro:*
7: em_estlro:*
8: em_hlc:*
9: gb_estl1:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrlc:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40.4	20.5	817	12	A2157865 SP_0059_A
2	39.8	20.2	1202	10	BE962592 601655935
3	39	19.8	266	12	A2493089 1M0327M20
4	39	19.8	568	12	A2271424 RPI-23-4
5	38.6	19.6	1344	10	BG111381 602281745
6	38.2	19.4	266	12	BH124550 RPI-24-2
7	38.2	19.4	1163	10	BR310265 601894769
8	37.8	19.2	389	9	AA818956 UI-R40-A
9	37.8	19.2	997	12	CNS0057E
10	37.6	19.1	421	9	AA667914 v19109.r
11	37.6	19.1	440	12	A2170914 SP_0118_A
12	37.6	19.1	477	12	AO881531 HS_5253_A
13	37.6	19.1	492	9	BB853900 BB853900
14	37.6	19.1	493	9	BB853873 BB853873
15	37.6	19.1	769	12	A2206648 SP_0105_B
16	37.4	19.0	412	10	BR143006 601791786
17	37.4	19.0	416	10	BE283447 601103455

C 18	37.4	19.0	1303	10	BG026974
C 19	37.2	18.9	399	12	AO821596
C 20	37.2	18.9	758	12	AG035520
C 21	36.8	18.7	320	9	AW198084
C 22	36.8	18.7	875	12	AG130893
C 23	36.8	18.7	1026	10	BG105107
C 24	36.8	18.7	1147	12	AO895354
C 25	36.8	18.7	2099	10	BG294288
C 26	36.6	18.6	415	10	BF062823
C 27	36.6	18.6	965	10	BG823762
C 28	36.4	18.5	689	12	AG056509
C 29	36.4	18.5	1249	10	BF036764
C 30	36.4	18.5	1341	10	BF795965
C 31	36.2	18.4	302	10	W01197
C 32	36.2	18.4	979	12	CNS070C1
C 33	36.2	18.4	1571	10	BF101099
C 34	36	18.3	227	12	AG024885
C 35	36	18.3	348	10	BG353106
C 36	36	18.3	451	9	BB852401
C 37	36	18.3	465	9	BB854380
C 38	36	18.3	564	12	A2630243
C 39	36	18.3	750	12	CNS01Y5V
C 40	36	18.3	967	10	BG353588
C 41	36	18.3	1001	12	AG107552
C 42	36	18.3	1024	10	BM415981
C 43	36	18.3	1219	10	BE966129
C 44	35.8	18.2	286	12	AQ103027
C 45	35.8	18.2	693	12	BH575352

ALIGNMENTS

RESULT 1
LOCUS A2157865
DEFINITION SP_0059_A1-B04-SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate-59 Col-7 Row-C, DNA sequence.
ACCESSION A2157865
VERSION A2157865.1 GI:8310468
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus.
REFERENCE Cameron,R.A., Mahalras,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.D., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehnach,H., Britten,R.J., Davidson,E.H. and Hood,L.
TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
MEDLINE 20402566
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 59 row: C column: 7
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 817.
Location/Qualifiers
1. 817
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="plate=59 Col=7 Row=C"
/clone_lib="Strongylocentrotus purpuratus, purple sea

```
urclin, sperm genomic BAC library"
/notes="Organ: sperm; Vector: BACs.6; BAC Clones in E-Coli
DH10B"
BASE COUNT      153 a      217 c      141 g      306 t
ORIGIN
```

	Query Match	20.5%	Score 40.4	DB 12	Length 817
	Best Local Similarity	61.3%	Pred. No. 31		
	Matches 65	Conservative 0	Mismatches 41	Indels 0	Gaps 0
Oy	1 gcatcgaagcgaaggtctgctctgcctctcgcgcacccatctctctctctgata 60				
Db	516 gctgacgctc 575				
Oy	61 ctctctctctctctctctctgatactcactcactctgcgcgtgcggttc 106				
Db	576 gctctgctc 621				

[illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	1 (bases 1 to 1207)			
1 (bases 1 to 1207)				
NIH-MGC http://mgc.ncl.nih.gov/				
National Institutes of Health, Mammalian Gene Collection (MGC)				
Unpublished (1999)				
On Oct 3, 2000 this sequence version replaced g1:10573297.				

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
Plate: L10K570 row: e column: 10
High quality sequence, stop: 55.

FEATURES	SOURCE	LOCATION/Qualifiers
1.	.1202	
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="IMAGE:3855729"		
/clone_1lb="NIR_MGC_66"		
/tissue_type="adenocarcinoma"		
/lab_host="DH10B (phage-resistant)"		
/note="Organ: ovary; Vector: PMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: oligo dm-Average insert size 1.8 kb. Library constructed by Life Technologies."		
BASE COUNT	654 a	101 c 366 g 61 t
ORIGIN		

	Query Match	20.2%	Score 39.8;	DB 10;	Length 1202;
	Best Local Similarity	64.8%	Pred. No. 42;		
	Matches	59;	Conservative	0;	Mismatches 32; Indels 0; Gaps 0;
OY	17	tttgcgtcttcctctcagatccacactcttcctctcttgctctcaactctctctcttctt	76		
Db	919	tttctttccctttttgacctttttctctctctctctctctctctctctctctctctt	920		
OY	77	tgtgatcctcacactctggcgtlpgccgattcc	107		

Db 919 TCTCCTCTTTTCCCTTGTCTTCCCTCTCC 889

	RESULT	3
	AZ493089/c	
	LOCUS	
	DEFINITION	266 bp DNA linear GSS 05-OCT-2000
	ACCESSION	AZ493089 Mouse 10kb plasmid UUCGIM library Mus musculus genomic clone UUCGIM0327M20 F, DNA sequence.
	VERSION	AZ493089
	KEYWORDS	AZ493089.1 GI:10666263
	SOURCE	G.S.
	ORGANISM	house mouse. Mus musculus

REFERENCE
AUTHORS
1 (bases 1 to 266)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddnn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0327 row: M column: 20
Seq primer: CGTGTGAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 266.

```

FEATURES
SOURCE
    location/Qualifiers
    1. 266
    /organism="Mus musculus"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="U06C1M0327M20"
    /clone_id="Mouse 10kb plasmid U06C1M library"
    /sex="Male"
    /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
    /note="Vector: PMD42nv; Purified genomic DNA from M.
    musculus C57BL/6J (male) was obtained from the Jackson
    Laboratory Mouse DNA Resource
    (http://www.jax.org/resources/documents/dnares/). The DNA
    was hydrodynamically sheared by repeated passage through a
    0.005 inch orifice at constant velocity. The sheared DNA
    was blunt end-repaired with T4 DNA polymerase and T4
    polynucleotide kinase. Adaptor oligonucleotides were
    ligated to the blunt ends in high molar excess. The
    adaptor DNA was purified and size-selected for a 9.5 to
    10.5 kb range using preparative agarose gel
    electrophoresis. Vector DNA was prepared from a derivative
    of PMD42 (g11473211419b/AP129072.1), a copy-number
    inducible derivative of plasmid R1. The vector was ligated
    with adaptors complementary to the insert adaptors and
    purified. The sheared, adaptor mouse DNA was annealed to
    adaptor vector DNA, and transformed into
    chemically-competent E. coli XL10-Gold (Stratagene) cells
    and selected for ampicillin resistance."

```

Query Match	19.8%;	Score 39;	DB 12;	Length 266;
Best Local Similarity	59.5%;	Pred. No. 65;		
Matches 66;	Conservative 0;	Mismatches 45;	Indels 0;	Gaps 0;

QY 22 tgccttcctcagcttcacatcttcctctgtctgctacatcttccttccttctgta 81
 DB 256 tccctccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 197
 QY 82 tctctcactctgcgctgcggtcccaagaagcagtgcccaatcactact 132
 DB 196 tttcttcttcttcttctcctccatctcattgacccagccgcttgact 146

RESULT 4
 LOCUS A2271424
 DEFINITION RPCI-23-444K19.TV RPCI-23 Mus musculus genomic clone RPCI-23-444K19
 DNA sequence.
 ACCESSION A2271424
 VERSION A2271424.1 GI:9485041
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 568)
 Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akiret,
 B., Leinus, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M. End Sequences from Library RPCI-23
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSSs: RPCI-23-444K19.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.buffalo.edu). Clones can be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/purchasingframe.htm>)
 or from Resea ch Genetics (http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html)
 Plate: 444 row: K column: 19
 Seq primer: 17
 Class: BAC ends.

FEATURES
 source
 1..568
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-444K19"
 /clone_1ib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1;
 EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 202 a 93 c 141 g 132 t
 ORIGIN

Query Match 19.8%; Score 39; DB 12; Length 568;
 Best Local Similarity 59.5%; Pred. No. 64;
 Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 22 tgccttcctcagcttcacatcttcctctgtctgctacatcttccttccttctgta 81
 DB 550 tccctccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 197
 QY 82 tctctcactctgcgctgcggtcccaagaagcagtgcccaatcactact 132

DB 490 tttcttcttcttcttctcctccatctcattgacccagccgcttgact 440

RESULT 5
 LOCUS BG111381
 DEFINITION 602281745F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4369259 5',
 mRNA sequence.
 ACCESSION BG111381
 VERSION BG111381.1 GI:12604803
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1344)
 NIH-MGC <http://nigc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ARCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LHAM10024 row: j column: 12
 High quality sequence stop: 520.

FEATURES
 source
 1..1344
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4369259"
 /clone_1ib="NIH-MGC-86"
 /tissue_type="osteosarcoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: bone; Vector: pCMV-Sport6; Site: 1; NotI;
 Site: 2; SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.533 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH-MGC Library."

BASE COUNT 446 a 396 c 393 g 109 t
 ORIGIN

Query Match 19.6%; Score 38.6; DB 10; Length 1344;
 Best Local Similarity 53.7%; Pred. No. 76;
 Matches 80; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 18 ttcgtcttcctcagcttcacatcttcctctgtctgctacatcttccttccttct 77
 DB 219 taccctctgctgctccgctcctcctcctcctcctcctcctcctcctcctt 160
 QY 78 gtaatctcactctgcgctgcggtcccaagaagcagtgcccaatcactactgcaac 137
 DB 159 ctctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 111
 QY 138 aggaagactggttaagtgcagagagatg 166
 DB 99 aggaagagagagagagagagagagagag 71

RESULT 6
 LOCUS BH124550
 DEFINITION RPCI-24-289F3.TV RPCI-24 Mus musculus genomic clone RPCI-24-289F3,
 DNA sequence.
 ACCESSION BH124550
 VERSION BH124550.1 GI:14968062
 KEYWORDS GSS.

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 266)
AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akintel, B., Levins, M., Tsagayev, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other GSS: RPCI-24-289F3.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 289 row: F column: 3
 Seq primer: T7
 Class: BAC ends
FEATURES Location/Qualifiers
 1..266
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-289F3"
 /clone_id="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pFARAPCI. Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pFARAPCI cloning vector at the BamHI sites using XhoI partially digested male C57BL/6J DNA."
BASE COUNT 47 a 87 c 18 g 114 t
ORIGIN
 Query Match 19.4%; Score 38.2; DB 12; Length 266;
 Best Local Similarity 73.1%; Pred. No. 96;
 Matches 49; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 Oy 26 ttcttcacaggttcacatctctctctgtgctacacctctctctctgtgacgc 85
 Db 171 tctctcttctcttctcttctctctctcttctcttctcttctcttctcttctctc 230
 Oy 86 tcaactc 92
 Db 231 tctctctc 237
RESULT 7
LOCUS BF310265/c 1163 bp mRNA linear EST 21-NOV-2000
DEFINITION 601894769F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4124249.5', mRNA sequence.
ACCESSION BF310265
VERSION BF310265.1 GI:11257806
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1163)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs.fremail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
 Plate: LCM1010 row: 1 column: 18
 High quality sequence stop: 506.
FEATURES Location/Qualifiers
 1..1163
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4124249"
 /clone_id="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally adaptor: GGGCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH-MGC Library."
BASE COUNT 325 a 297 c 348 g 192 t
ORIGIN
 Query Match 19.4%; Score 38.2; DB 10; Length 1163;
 Best Local Similarity 63.7%; Pred. No. 93;
 Matches 58; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 Oy 9 gccacgagcttcgctctctctcagcttcacatctctctgtgctacacctctc 68
 Db 1058 gcttgcgcttcttctcttcttcttcttcttcttcttcttcttcttcttcttctt 999
 Oy 65 ctcttcctgtgctatctcactcctctgcgctg 99
 Db 998 ctctgcctctgcttctctctctctctctctctc 968
RESULT 8
LOCUS AA818956 389 bp mRNA linear EST 03-JUL-1999
DEFINITION UI-R-A0-as-e-11-0-UI-s1 UI-R-A0 Rattus norvegicus cDNA clone
 UI-R-A0-as-e-11-0-UI-3, similar to gblAF001308/AF001308
 Arabidopsis thaliana chromosome IV BAC clone T10M13, complete
 sequence (Arabidopsis thaliana), mRNA sequence.
ACCESSION AA818956
VERSION AA818956.1 GI:4227624
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 389)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Feb 17, 1998 this sequence version replaced g1:2888542.
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: mscare@blue.weeg.iowa.edu
 The sequence tag present in the cDNA between the NotI site and the

oligo-dr track served to identify it as a clone from the normalized adult Brain library. cDNA library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE ID-1767986 The following repetitive elements were found in this cDNA sequence: 1-92, >POLY-A#simple_repeat 310-389, >B#SINE/B2 Seq primer: M13 Forward

FEATURES

source

Location/Qualifiers

1. 389
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UT-R-A0-as-e-11-0-UT"
/clone_lib="UT-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="vector: pTR73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; This library consists of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 86 a 82 c 74 g 147 t

ORIGIN

Query Match 19.2%; Score 37.8; DB 9; Length 389;
Best Local Similarity 53.8%; Pred. No. 1.2e+02;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Oy 17 ttctggtcttcttcagcttcacatcttctcttctgctacatcttcttct 76
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 29 ttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 88
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 77 tctgctctcactctctgcgcgttcctcaagaagccagtcacatctactgcaa 136
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89 tcttccaccatcatcatcatcatcatcatcatcatcatcatcatcatcat 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 137 caggagagactgtaagtcagagag 161
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149 CAAGACAGAGTTCGATCCGACGAG 173
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9

CNS005TE

LOCUS

DEFINITION

CNS005TE 997 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence JET3 end of BAC #
BACR1K22 of RPEC-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL060767
VERSION AL060767.1 GI:4943573
KEYWORDS GSS.

SOURCE

ORGANISM

fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 997)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and

Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPEC-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

Location/Qualifiers

1. 997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPEC-98"
/clone="BACR1K22"
/note="end : TET3"
BASE COUNT 89 a 99 c 13 g 258 t 538 others

ORIGIN

Query Match 19.2%; Score 37.8; DB 12; Length 997;
Best Local Similarity 36.4%; Pred. No. 1.1e+02;
Matches 28; Conservative 33; Mismatches 16; Indels 0; Gaps 0;

Oy 17 ttctggtcttcttcagcttcacatcttctcttctgctacatcttcttct 76
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 818 tcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 877
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 77 tctgctctcactctctt 93
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 878 tcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 894
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10

AA667914

LOCUS

DEFINITION

AA667914 421 bp mRNA linear EST 19-NOV-1997
v019f09.r1 StrataGene mouse heart (#937316) Mus musculus cDNA clone
IMAGE:1222889 5', mRNA sequence.

ACCESSION AA667914
VERSION AA667914.1 GI:2626615
KEYWORDS EST.

SOURCE

ORGANISM

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 421)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellander, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

COMMENT

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
WGI:648481
Seq primer: -28m3 rev1 ET from Amersham
High quality sequence stop: 399.
Location/Qualifiers
1. 421
/organism="Mus musculus"
/strain="N1H/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:1222889"
/clone_lib="StrataGene mouse heart (#937316)"
/sex="pooled"

```

/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT      70 a      139 c      67 g      145 t
ORIGIN

Query Match      19.1%; Score 37.6; DB 9; Length 421;
Best Local Similarity 59.3%; Pred. No. 1.3e+02;
Matches 64; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 1 gcatgcagcagatgggttcgtcttcctcagctccatctctctctgtgctca 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 gTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTT 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 61 ctctctctctctctctctctctctctctctctctctctctctctctctca 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 CCCATCTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
A2170914      440 bp      DNA      linear      GSS 29-AUG-2000
LOCUS
DEFINITION
SP_0118_1_802_77A Strongylocentrotus purpuratus, purple sea urchin
clone Plate-118 Col-3 Row-C, DNA sequence.
A2170914
VERSION
A2170914.1 GI:8341282
KEYWORDS
GSS.
ORGANISM
Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE
1 (bases 1 to 440)
Cameron,R.A., Mahalras,G., Rast,J.P., Martinez,P., Blondi,T.R.,
Swartzell,S., Wallace,J.C., Pouška,A.J., Livingston,B.T., Wray,
G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
20402565
CONTACT: Cameron, RA, Davidson, EH, Hood, L
DIVISION OF BIOLOGY 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 118 row: C column: 3
Seq primer: 77
Classes: BAC ends
High quality sequence stop: 440.
Location/Qualifiers
1..440
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate-118 Col-3 Row-C"
/clone_id="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli
DH10B"

BASE COUNT      85 a      124 c      60 g      168 t      3 others
ORIGIN

```

```

Query Match      19.1%; Score 37.6; DB 12; Length 440;
Best Local Similarity 72.1%; Pred. No. 1.3e+02;
Matches 49; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 10 ccatgggttcgtctctctcagctccatctctctctctctctctctctc 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 CCCCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCC 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 70 ttctctct 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 TTTTCTT 382

RESULT 12
AO881531/c      477 bp      DNA      linear      GSS 09-NOV-1999
LOCUS
DEFINITION
HS_5253_A1_F03_47A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-9021 Col-5 Row-K, DNA sequence.
AO881531
VERSION
AO881531.1 GI:6312998
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 477)
Mahalras,G.G., Wallace,J.C., Smith,R., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
CONTACT: Mahalras GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3687
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 9021 row: K column: 5
Seq primer: 77
Classes: BAC ends
High quality sequence stop: 477.
Location/Qualifiers
1..477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-9021 Col-5 Row-K"
/clone_id="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"

BASE COUNT      273 a      8 c      115 g      26 t      55 others
ORIGIN

Query Match      19.1%; Score 37.6; DB 12; Length 477;
Best Local Similarity 56.8%; Pred. No. 1.3e+02;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 17 ttctgcctctctcagctccatctctctctctctctctctctctctctc 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 TTTCCTATTCTCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTTC 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

? MOLECULE TYPE: Genomic DNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? FRAGMENT TYPE: <unknown>
? ORIGINAL SOURCE:
? SEQUENCE DESCRIPTION: SEQ ID NO: 17
US-10-125-767-17

```

Query Match	17.3%	Score 34;	DB 6;	length 42399;
Best Local Similarity	57.5%;	Pred. NO. 4.9;		
Matches 61;	Conservative 0;	Mismatches 45;	Indels 0;	Gaps 0;

OY 2 cgtcgaacccatgagcttcgcgtccttcccctcagtccatctctcctcttggtctaac 61
| | | | | | | | | | | | | | | |
Db 41342 CTTCCTCTCCCTGTTTCTTCCTCCTTCTCTTCCTTTCTCTCCTTCCTTTAGGTATP 41401

QY	62 cttcttcttcttccttgatcctccactctgcgcgtgcggtcca	107
Db	41402 TCTTCTCTCTTCCTTTCCCTTTCTCTCTCTTCTGTGCTTTTTC	41447

```

RESULT      2
US-10-105-299-1441
: Sequence 1441, Application US/10105299
: GENERAL INFORMATION:
:   APPLICANT: Rosen, et. al
:   TITLE OF INVENTION: Human Secreted Proteins
:   FILE REFERENCE: PS950
:   CURRENT APPLICATION NUMBER: US/10/105,299
:   CURRENT FILING DATE: 2002-03-26
:   NUMBER OF SEQ ID NOS: 15197
:   Prior Application removed - See File Wrapper or Palm
:   SOFTWARE: PatentIn Ver. 2.0
:   SEQ ID NO 1441
:   LENGTH: 1377
:   TYPE: DNA
:   ORGANISM: Homo sapiens
: US-10-105-299-1441

```

Query Match	16.9%;	Score 33.2;	DB 6;	Length 1377;
Best Local Similarity	64.1%;	Pred. No. 4.3;		
Matches 50;	Conservative 0;	Mismatches 28;	Indels 0;	Gaps 0;

```

Oy 18 ttcgtgcctctctctcagccttcacatcttcctctctgtgtctacctctctctcttcctt 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 835 ttcctctctctctctctctctctctctctctctctctctctctctctctctctctct 894

```

```
Oy      78  gtgatctctcaacctgc 95
          |  |||||  |||||  |
Db      895  ctctctctctctctcc 912
```

```

RESULT      3
US-09-782-959A-17
: Sequence 17, Application US/09782959A
: GENERAL INFORMATION:
: APPLICANT: SADEE, MOLEFANG
: APPLICANT: BOTKA, CHRISTOPHER
: TITLE OF INVENTION: NOVEL MEMBERS OF THE H+/OLIGOPEPTIDE TRANSPORTER GENE FAMILY
: FILE REFERENCE: 305t-001010US
: CURRENT APPLICATION NUMBER: US/09/782,959A
: PRIOR FILING DATE: 2002-03-26
: PRIOR APPLICATION NUMBER: US 60/182,328
: PRIOR FILING DATE: 2000-02-14
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO: 17
: LENGTH: 4182
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-782-959A-17

```

Query Match	16.9%	Score 33.2	DB 5	Length 4182
Best Local Similarity	67.1%	Pred. No. 5.3		
Matches 47; Conservative	0	Mismatches 23	Indels 0	Gaps 0

QY 26 ttctctcagcttcacatctctctcttggtgtactctctctctcttccttgatctc 85
 ||||| ||| ||||||||| | |||| | || | ||||
 Db 1557 ttc 1616

QY	86	tcactctctgc	95
Db	1617	tcctctcttcc	1626

RESULT 4
PCT-US02-09785-1053/c
; Sequence 1053, Application PC/TUS0209785

```

1  APPLICANT: Human Genome Sciences, Inc.
2  TITLE OF INVENTION: Human Secreted Protein
3  FILE REFERENCE: PS095PCP
4  CURRENT APPLICATION NUMBER: PCT/US02/09785
5  CURRENT FILING DATE: 2002-03-19
6  PRIOR APPLICATION NUMBER: US 60/331,287
7  PRIOR FILING DATE: 2001-11-13
8  PRIOR APPLICATION NUMBER: US 60/306,171
9  PRIOR FILING DATE: 2001-07-19
10 PRIOR APPLICATION NUMBER: US 60/277,340
11 PRIOR FILING DATE: 2001-03-21
12 NUMBER OF SEQ ID NOS: 1130
13 SOFTWARE: PatentIn Ver. 2.0
14 SEQ ID NO 1053
15 LENGTH: 5754
16 TYPE: DNA
17 ORGANISM: Homo sapiens
18 PCT-US02-09785-1053

```

Query Match:	16.98:	Score 33.2:	DB 1:	length 5754:
Best Local:	67.18:	Pred. No. 5.6:		
Matches 47:	Conservative	0:	Mismatches 23:	Indels 0:
				Gaps 0:

Qy	26 tttctctcagactccatcttctctctctgtgtagtaacacctctctcttctctcttgtagtaccc 85 .
Db	2728 ttctcttccctttttcttctctctttcttttccttttctttctttctttctttctttctttc 2669

QY	86	tcactcttcg	95
		11111111	
Db	2668	TCTCTCTTTC	2659

```

RESULT 5
US-10-105-299-13322/c
: Sequence 13322, Application US/10105299
: GENERAL INFORMATION:
: APPLICANT: Rosen, et. al
: TITLE OF INVENTION: Human Secreted Proteins
: FILE REFERENCE: PS950
: CURRENT APPLICATION NUMBER: US/10/105,299
: CURRENT FILING DATE: 2002-03-26
: NUMBER OF SEQ ID NOS: 15197
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 13322
: LENGTH: 5754
: Type: DNA
: ORGANISM: Homo sapiens
US-10-105-299-13322

```

Query Match	Score	DB	Length
Best Local Similarity	67.18;	Pred. No. 5.6;	
	16.98;		5754;
		DB 6;	
	33.2;		

Matches	47; Conservative	0; Mismatches	23; Indels	0; Gaps	0; Ambiguous
Qy	26	ttttctcagactccatcttctcctatctgtagtgcactcttctctctcttgatgac	85		
Db	2728	ttttctttccctttttcttctctctctctctctctctctctctctctctctctctc	2669		
Qy	86	tcactctctgc	95		
Db	2668	tcctctcttttc	2659		

```

1 RESULT 6
2 PCT-US02-09785-1058/c
3 Sequence 1058, Application PCT/TUS0209785
4 GENERAL INFORMATION:
5 APPLICANT: Human Genome Sciences, Inc.
6 TITLE OF INVENTION: Human Secreted Proteins
7 FILE REFERENCE: PS905PCT
8 CURRENT APPLICATION NUMBER: PCT/US02/09785
9 PRIOR FILING DATE: 2002-03-19
10 PRIOR APPLICATION NUMBER: US 60/331,287
11 PRIOR FILING DATE: 2001-11-13
12 PRIOR APPLICATION NUMBER: US 60/306,171
13 PRIOR FILING DATE: 2001-07-19
14 PRIOR APPLICATION NUMBER: US 60/277,340
15 PRIOR FILING DATE: 2001-03-21
16 NUMBER OF SEQ ID NOS: 1130
17 SOFTWARE: PatentIn Ver. 2.0
18 SEQ ID NO 1058
19 LENGTH: 5755
20 TYPE: DNA
21 ORGANISM: Homo sapiens
22 PCT-US02-09785-1058

```

Query Match	Best Local Similarity	16.9%	Score 33.2	DB 1	Length 5755
Matches	47	Conservative	0	Mismatches	23
				Indels	0
				Gaps	0
QY	26	tttctctcagcttcacatcttccctctctgtgtcctcctctctcttcttcttgatc	85		
Db	2730	ttttcttttttcttt	1111		
QY	86	tcaactctgc	95		
Db	2670	ttctctttttc	2661		

```

RESULT 7
US-10-105-299-13331/C
: Sequence 13331. Application US/10105299
: GENERAL INFORMATION:
:   APPLICANT: Rosen, et al
:   TITLE OF INVENTION: Human Secreted Proteins
:   FILE REFERENCE: P8950
:   CURRENT APPLICATION NUMBER: US/10/105,299
:   CURRENT FILING DATE: 2002-03-26
:   NUMBER OF SEQ ID NOS: 1519
:   Prior Application removed - See File Wrapper or Palm
:   SOFTWARE: PatentIn Ver. 2.0
:   SEQ ID NO 13331
:   LENGTH: 5755
:   TYPE: DNA
:   ORGANISM: Homo sapiens
US-10-105-299-13331

```

```
Query Match      16.98; Score 33.2; DB 6; Length 5755;
Best Local Similarly 67.1%; Pred. No. 5.6;
Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0

QY      26 ttcttcagacttcaactttccctctctgtggtcctaacttctctcttcccttgatcacgc 85
||||| | | | | | | | | | | | | | | | | | | |
```

D_b 2730 TTTCTTTCCTTTTTCCTCCTTCTGTTCTTCTCTTCTTCTTCTCTGC 2672
 O_Y 86 tcaactcttcgc 95
 || ||||| |
 D_b 2670 TCCTCTCTTC 2661

```

RESULT 8
PCT-US02-01109-32
: Sequence 32, Application PC/TUS0201109
: GENERAL INFORMATION:
: APPLICANT: Human Genome Sciences, Inc.
: TITLE OF INVENTION: 50 Human Secreted Proteins
: FILE REFERENCE: PZ016PCT2
: CURRENT APPLICATION NUMBER: PCT/US02/01109
: CURRENT FILING DATE: 2002-01-17
: PRIOR APPLICATION NUMBER: US 60/262,066
: PRIOR FILING DATE: 2001-01-18
: NUMBER OF SEQ. ID NOS: 206
: SOFTWARE: PatentIn Ver. 2.0
: SEQ. ID NO 32
: LENGTH: 626
: TYPE: DNA
: ORGANISM: Homo sapiens
PCT-US02-01109-32

```

[illegible]

```

RESULT      9
US-10-114-893-128
: Sequence 128. Application US/10114893
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John M.
: APPLICANT: Lavallee, Edward R.
: APPLICANT: Collins-Racle, Lisa A.
: APPLICANT: Evans, Cheryl
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Bowman, Michael R.
: APPLICANT: Spaulding, Vikki
: APPLICANT: Callin-Duckett, McKeough
: APPLICANT: Kelleher, Kerry S.
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
: FILE REFERENCE: GI 6000-10A
: CURRENT APPLICATION NUMBER: US/10/114,893
: CURRENT FILING DATE: 2002-04-02
: EARLIER APPLICATION NUMBER: 09/413,232
: EARLIER FILING DATE: 1999-10-06
: NUMBER OF SEQ ID NOS: 321
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 128
: LENGTH: 1199
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-114-893-128

```

Query Match	16.8%;	Score 33;	DB 6;	Length 1199;
Best Local Similarity	60.7%;	Pred. No. 4.8;		

? PRIOR APPLICATION NUMBER: US 09/950,083
 ? PRIOR FILING DATE: 2001-09-12
 ? NUMBER OF SEQ. ID NOS: 1834
 ? SOFTWARE: patentIn Ver. 2.0
 ? SEQ. ID NO. 909
 ? LENGTH: 9718
 ? TYPE: DNA
 ? ORGANISM: Homo sapiens
 ? CCT-0502-09370-909

Query Match	15.9%;	Score 31.4;	DB 1;	Length 9718;
Best Local Similarity	48.1%;	Pred. No. 18;		
Matches	89;	Conservative	0;	Mismatches 96;
				Indels 0;
				Gaps 0;

QY	4	tcgagcgaatggatttcgagcttctctcagttccatcattccattctctgtgtctactc	63
Db	7075	TCCTCTCCTCCCTGGGTAAGCTGGAGACTGAGGACAGCTGCCACCATGCTCGCTGGTTTATT	7016
QY	64	tctctcttctctcttgtaetctctcaactctctgcgctgcggttccaaagaacgaatgc	123
Db	7015	TTTTTGTTTTTTAGAGGCTCTCACTATGTTGGCCAGGCTGCTCAACACTTTGGGCTCC	6956
QY	124	tcatctactgcacacgaggaactcgtlaaagtcacagaaagatcgtagctcgcgcgaac	183
Db	6955	ACCAATCTTCCCGCCCTCAGACCTTCCGTGATGCTGTAAAGTCTGGGATTAACAGGTC	6886

QY	184	gtgtc	188
Db	6895	GTGCC	6891

```

RESULT 14
US-10-105-299-6278/c
: Sequence 6278, Application US/10105299
: GENERAL INFORMATION:
: APPLICANT: Rosen, et. al
: TITLE OF INVENTION: Human Secreted Proteins
: FILE REFERENCE: PS950
: CURRENT APPLICATION NUMBER: US/10/105,299
: CURRENT FILING DATE: 2002-03-26
: NUMBER OF SEQ ID NOS: 15197
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 6278
: LENGTH: 9718
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-105-299-6278

```

Query Match	15.9%	Score 31.4	DB 6	Length 9718
Best Local Similarity	48.1%	Pred. No. 18		
Matches 89; Conservative	0	Mismatches 96	Indels 0	Gaps 0

[illegible]

OY	184	gtgtc	188
Db	6895	GTGCC	6891

RESULT 15

PCT-US02-09105-422/c
Sequence 422, Application PC/TUS0209105
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: P5915PCT
CURRENT APPLICATION NUMBER: PCT/US02/09105
CURRENT FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: US 60/278,650
PRIOR FILING DATE: 2001-03-27
PRIOR APPLICATION NUMBER: US 09/950,082
PRIOR FILING DATE: 2001-09-12
PRIOR APPLICATION NUMBER: US 09/950,083
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 779
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 422
LENGTH: 9728
TYPE: DNA
ORGANISM: Homo sapiens
PCT-US02-09105-422

Query Match	15.9%;	Score 31.4;	DB 1;	Length 9728;
Best Local Similarity	48.1%;	Pred. No. 18;		
Matches	89;	Mismatches	96;	Indels 0;
		Conservative		Gaps 0;

[illegible]

Search completed: May 12, 2002, 21:35:57
Job time: 6853 sec

Search completed: May 12, 2002, 21:35:57
Job time: 6853 sec

Query Match
Best Local Similarity 100.0%; Score 197; DB 6; Length 197;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 32 a 56 c 49 g 60 t

ORIGIN

/protein_id="CA870495.1"
/db_xref="GI:6779835"
/translation="MGFVLSQPSFLVSTLLFLVISHSCRAGSKKPVPIICNRR
TGKQHM"

Query Match
Best Local Similarity 100.0%; Score 197; DB 6; Length 197;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 32 a 56 c 49 g 60 t

ORIGIN

/protein_id="CA870495.1"
/db_xref="GI:6779835"
/translation="MGFVLSQPSFLVSTLLFLVISHSCRAGSKKPVPIICNRR
TGKQHM"

RESULT 2
LOCUS A98895 211 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 3 from Patent WO9509189.
ACCESSION A98895
VERSION A98895.1 GI:6781856
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 211)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR ANDROCTONINE, VECTOR CONTAINING SAME AND
TRANSFORMED DISEASE-RESISTANT PLANTS OBTAINED
JOURNAL Patent: WO 9509189-A 3 25-FEB-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source location/Qualifiers
1..211
CDS
/organism="unidentified"
/db_xref="taxon:32644"
/note="unnamed protein product"
/codon_start=1
/protein_id="CA870573.1"
/db_xref="GI:6781857"
/translation="MGFVLSQPSFLVSTLLFLVISHSCRARSVCROIKICRRRG
GCTVCKTNRPY"

BASE COUNT 33 a 58 c 57 g 63 t

ORIGIN

Query Match
Best Local Similarity 77.0%; Score 117.4; DB 6; Length 211;
Matches 161; Conservative 0; Mismatches 36; Indels 12; Gaps 1;

BASE COUNT 33 a 58 c 57 g 63 t

ORIGIN

Db 121 TCACAGATCTGCAGAGAGAGGGGTGCTCTACTACAGTGCATACAGCCACTACTAG 180
Qy 169 CTCGCGCAGGCGAAGCTGTCCAGCATCC 197
Db 181 CTCGCGCAGGCGAAGCTGTCCAGCATCC 209

RESULT 3
LOCUS A95926 106 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 4 from Patent WO924594.
ACCESSION A95926
VERSION A95926.1 GI:6779832
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 106)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR ANDROCTONINE, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 924594-A 4 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source location/Qualifiers
1..106
CDS
/organism="unidentified"
/db_xref="taxon:32644"
/note="unnamed protein product"
/codon_start=1
/protein_id="CA870494.1"
/db_xref="GI:6779833"
/translation="MGFVLSQPSFLVSTLLFLVISHSCRA"

BASE COUNT 8 a 34 c 20 g 44 t

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 103; DB 6; Length 106;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 8 a 34 c 20 g 44 t

ORIGIN

RESULT 4
LOCUS A98894 106 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 2 from Patent WO9509189.
ACCESSION A98894
VERSION A98894.1 GI:6781854
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 106)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR ANDROCTONINE, VECTOR CONTAINING SAME AND
TRANSFORMED DISEASE-RESISTANT PLANTS OBTAINED
JOURNAL Patent: WO 9509189-A 2 25-FEB-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source location/Qualifiers
1..106
CDS
/organism="unidentified"
/db_xref="taxon:32644"
/note="unnamed protein product"


```

/codon_start=1
/protein_id="CA870572.1"
/db_xref="GI:6781835"
/translation="MGFVLPSPSPFLVSTLLFLVISHSCRA"
BASE COUNT      8 a      34 c      20 g      44 t
ORIGIN

```

Query Match	52.3%	Score 103;	DB 6;	Length 106;
Best Local Similarity	100.0%	Pred. No. 1.1e-20;		
Matches 103; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

RESULT	5
AX014749	
LOCUS	AX014749 120 bp DNA linear PAT 07-SEP-2000
DEFINITION	Sequence 4 from Patent W0953053.
ACCESSION	AX014749
VERSION	AX014749.1 GI:10041017
KEYWORDS	.
SOURCE	synthetic construct.
ORGANISM	artificial sequence.
REFERENCE	1 (bases 1 to 120)
AUTHORS	Lamberty,M., Hofmann,J., Bulet,P. and Brookhart,G.L.
TITLE	Gene coding for heliomycin and use thereof
JOURNAL	Patent: WO 9553053-A 4 21-OCT-1999;
	LABERRY MIRELIEE (FR); HOFMANN JULES (FR); BULET PHILIPPE (FR);
	RHONE POULENC AGROCHIMIE (FR); BROOKHART GARY LEE (US)
FEATURES	Location/Qualifiers

Query Match	51.6%;	Score 101.6;	DB 6;	Length 120;
Best Local Similarity	92.2%;	Pred. NO. 2.9e-20;		
Matches 107; Conservative	0;	Mismatches 9;	Indels 0;	Gaps 0

RESULT	6			
A95925				
LOCUS	A95925	98 bp	DNA	
DEFINITION	Sequence 3 from Patent WO924594.			
ACCESSION	A95925			
VERSION	A95925.1	GI:6779830		
KEYWORDS				
SOURCE	unidentified.			

ORGANISM	REFERENCE
unidentified unclassified	1 (bases 1 to 98)
DEROSE, R.	HOFFMANN, J.
AUTHORS	GENE CODING FOR THAMNAIN, VECTOR CONTAINING SAME AND RESULTING
TITLE	TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL	PATENT: WO 9924594-A 3 20-MAY-1999.
DEROSE RICHARD (FR);	HOFFMANN JULES (FR)

Query Match	48.7%	Score 96:	DB 6:	Length 98:
Best Local Similarity	100.0%	Pred. No.	1.4e-18:	
Matches 96:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:

RESULT	7			
LOCUS	AX014748	261 bp	DNA	Linear
DEFINITION	Sequence 3 from Patent WO953053.			
ACCESSION	AX014748			
VERSION	AX014748.1	GI:10041015		
KEYWORDS				
SOURCE	synthetic construct.			
ORGANISM	synthetic construct			
REFERENCE	artificial sequence.			
AUTHORS	1 (bases 1 to 261)			
TITLE	Lambertzy,M., Hofmann,J., Bulet,P. and Brookhart,G.L.			
JOURNAL	Gene coding for helminthocline and use thereof			
	Patent: WO 953053-A 3 21-OCT-1999;			

Query Match	47.4%	Score 93.4	DB 6	length 261
Best Local Similarity	98.9%	Pred. No. 9.5e-18		
Matches 94	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	10	ccatgggtttcggtcttctctcagctccatctctctcttcttcttcttcttcttcttc	69	

Db 1 CCATGGCTTTCCTTTCTCTCAGCTTCACATCTTTCCTTGTGCTACTCTCTTC 60
OY 70 tttctcttgatctctcaactctgacgagc 104
|||||
Db 61 TTTCTGTGATCTCTCAGCTCTGCGCGCAT 95

RESULT 8

LOCUS A95931/c 87 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 9 from Patent WO924594.
ACCESSION A95931
VERSION A95931.1 GI:6779839
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 87)
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS
PATENT: WO 924594-A 9 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source 1..87
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 14 a 31 c 21 g 21 t
ORIGIN

Query Match 43.1%; Score 85; DB 6; Length 87;
Best Local Similarity 100.0%; Pred. No. 3.1e-15;

Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 113 gccagtgccaatcatctactgcacagagagactgtaagtccagagagatgtagctcg 172
|||||
Db 87 GCCAGTGCATCATCTACTGTCACACAGAGACTGTGAAGTCCAGAGATGTAGCTCG 28
OY 173 gcgagggcaacgctgcgaagatcc 197
|||||
Db 27 GCGAGCGCAACGTGTGACGCGATCC 3

RESULT 9

LOCUS A95928 75 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 6 from Patent WO924594.
ACCESSION A95928
VERSION A95928.1 GI:6779836
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 75)
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS
PATENT: WO 924594-A 6 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source 1..75
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 23 c 13 g 34 t
ORIGIN

Query Match

Best Local Similarity 98.7%; Pred. No. 1e-11;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gcgcgcagcagtggttcgtctctctcagctccatcttctctctgtgtctta 60

Db 1 GCGTCAGCGAGTGGCTTTCCTTCAGCTTCACATCTTTCCTTGTGCTACTCTCTTC 60
OY 61 ctctctctcttc 75
|||||
Db 61 CTCTCTCTCTTTC 75

RESULT 10

LOCUS A98896 75 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 4 from Patent WO9909189.
ACCESSION A98896
VERSION A98896.1 GI:6781858
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 75)
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR ANDROCTONINE, VECTOR CONTAINING SAME AND
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS OBTAINED
PATENT: WO 9909189-A 4 25-FEB-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source 1..75
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 23 c 13 g 34 t
ORIGIN

Query Match 37.3%; Score 73.4; DB 6; Length 75;
Best Local Similarity 98.7%; Pred. No. 1e-11;

Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 gcgtcgacgcgagtggttcgtctctctcagctccatcttctctctgtgtcta 60
|||||
Db 1 GCGTCGACGCGAGTGGCTTTCCTTCAGCTTCACATCTTTCCTTGTGCTACTCTCTTC 60
OY 61 ctctctctcttc 75
|||||
Db 61 CTCTCTCTCTTTC 75

RESULT 11

LOCUS AX014750 75 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 5 from Patent WO9953053.
ACCESSION AX014750
VERSION AX014750.1 GI:10041019
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 75)
AUTHORS Lambery,M., Hofmann,J., Bulet,P. and Brookhart,G.L.
TITLE Gene coding for helioleptin and use thereof
JOURNAL LAMBERY MIREILLE (FR); HOFFMANN JULES (FR); BULET PHILIPPE (FR);
RHONE POULENC AGRICHIMIE (FR); BROOKHART GARY LEE (US)
FEATURES
source 1..75
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 5 a 23 c 13 g 34 t
ORIGIN

Query Match

Best Local Similarity 98.7%; Pred. No. 1e-11;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gcgtcgaacgcatgggttcgtcttctcctcagcttcacatcttcctctgtgtctta 60
|||||
Db 1 GCGTCGACCGCGAGGGCTTCGTCTTTCCTCAGCTTCACCTTCCTTCGTGTCTA 60
OY 61 cctctctctcttcc 75
|||||
Db 61 CTCTCTCTTCTTCTTC 75

RESULT 12

A95929/c 72 bp DNA linear PAT 26-JAN-2000
LOCUS A95929 Sequence 7 from Patent WO9924554.
DEFINITION A95929
ACCESSION A95929.1 GI:6779837
VERSION A95929.1 GI:6779837
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 72)
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS
PATENT: WO 9924594-A 7 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source 1..72
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 33 a 11 c 23 g 5 t
ORIGIN

Query Match 34.2%; Score 67.4; DB 6; Length 72;
Best Local Similarity 98.6%; Pred. No. 6.6e-10;
Matches 68; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 35 gctccacatcttctctctgtgtctactcttctcttctcttctgtatctcactctg 94
|||||
Db 72 GCTTCACATCTTCTTCCTTCGTGTCTACTCTTCCTTCGTGTCTACTCTCTTG 13
OY 95 ccgtggccgg 103
|||||
Db 12 CCGTCCCG 4

RESULT 13

A98897/c 72 bp DNA linear PAT 26-JAN-2000
LOCUS A98897 Sequence 5 from Patent WO9909189.
DEFINITION A98897
ACCESSION A98897.1 GI:6781859
VERSION A98897.1 GI:6781859
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 72)
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR ANDROCTONINE, VECTOR CONTAINING SAME AND
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS OBTAINED
PATENT: WO 9909189-A 5 25-FEB-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source 1..72
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 33 a 11 c 23 g 5 t
ORIGIN

Query Match 34.2%; Score 67.4; DB 6; Length 72;
Best Local Similarity 98.6%; Pred. No. 6.6e-10;
Matches 68; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 35 gctccacatcttctctctgtgtctactcttctcttctcttctgtgatctcactctg 94
|||||
Db 72 GCTTCACATCTTCTTCCTTCGTGTCTACTCTTCCTTCGTGTCTACTCTCTTG 13
OY 95 ccgtggccgg 103
|||||
Db 12 CCGTCCCG 4

RESULT 14

AX014751/c 72 bp DNA linear PAT 07-SEP-2000
LOCUS AX014751 Sequence 6 from Patent WO9953053.
DEFINITION AX014751
ACCESSION AX014751
VERSION AX014751.1 GI:10041020
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 72)
AUTHORS Lamberly,M., Hoffmann,J., Bulet,P. and Brookhart,G.L.
TITLE Gene coding for heliothine and use thereof
JOURNAL PATENT: WO 9953053-A 6 21-OCT-1999;
LAMBERTY MIREILLE (FR); HOFFMANN JULES (FR); BULET PHILIPPE (FR);
PHONE POULENC AGROCHIMIE (FR); BROOKHART GARY LEE (US)
FEATURES
source 1..72
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 33 a 11 c 23 g 5 t
ORIGIN

Query Match 34.2%; Score 67.4; DB 6; Length 72;
Best Local Similarity 98.6%; Pred. No. 6.6e-10;
Matches 68; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 35 gctccacatcttctctctgtgtctactcttctcttctcttctgtatctcactctg 94
|||||
Db 72 GCTTCACATCTTCTTCCTTCGTGTCTACTCTTCCTTCGTGTCTACTCTCTTG 13
OY 95 ccgtggccgg 103
|||||
Db 12 CCGTCCCG 4

RESULT 15

A95924 63 bp DNA linear PAT 26-JAN-2000
LOCUS A95924 Sequence 2 from Patent WO9924594.
DEFINITION A95924
ACCESSION A95924
VERSION A95924.1 GI:6779828
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 63)
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS
PATENT: WO 9924594-A 2 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source 1..63
/organism="unidentified"
/db_xref="taxon:32644"
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70492.1"
/db_xref="GI:6779828"
/translation="GSKRPVPIYCNRRGKQRM"

BASE COUNT	19 a	14 c	19 g	11 t
ORIGIN	-			.

Query Match	32.08;	Score 63;	DB 6;	Length 63;
Best Local Similarity	100.00;	Pred. No. 1.4e-08;		
Matches 63; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

Search completed: May 12, 2002, 20:13:54
Job time: 7216 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 18:58:33 ; Search time 333.08 Seconds
(without alignments) 170.104 Million cell updates/sec

Title: US-09-554-024-1

Perfect score: 33
Sequence: 1 atccactactgcacagagagactgtaagtc 33

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1980.DAT.*
2:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1981.DAT.*
3:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1982.DAT.*
4:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1983.DAT.*
5:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1984.DAT.*
6:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1985.DAT.*
7:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1986.DAT.*
8:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1987.DAT.*
9:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1988.DAT.*
10:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1989.DAT.*
11:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1990.DAT.*
12:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1991.DAT.*
13:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1992.DAT.*
14:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1993.DAT.*
15:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1994.DAT.*
16:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1995.DAT.*
17:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1996.DAT.*
18:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1997.DAT.*
19:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1998.DAT.*
20:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA1999.DAT.*
21:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA2000.DAT.*
22:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA2001A.DAT.*
23:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA2001B.DAT.*
24:	/SIDSI/gcgdata/hold-geneseq/geneqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	33	100.0	33	20	AAK59770	CDNA encoding a th
2	33	100.0	63	20	AAK59771	CDNA encoding a th
3	33	100.0	98	20	AAK59772	CDNA encoding the
4	33	100.0	197	20	AAK59774	CDNA encoding a th
5	22.4	67.9	3102	17	AAH15600	laccase-1C3 gene.
6	20.2	61.2	589	22	AAH87754	Peppermint plant c
c	20.2	61.2	18073	20	AAH83948	Bacterial artifice
8	20	60.6	1275	23	AAH82520	DNA encoding novel
9	20	60.6	1737	23	AAH82525	DNA encoding novel

10	20	60.6	2289	23	AAH89143	DNA encoding novel
11	19.8	60.0	179	23	AAH81172	Enterococcus faeca
12	19.6	59.4	404	21	AAK26094	Human secreted pro
13	19.4	58.8	792	23	ABL21645	Drosophila melanog
14	19.4	58.8	1488	22	AAH83638	Novel human transp
15	19.4	58.8	1626	22	AAH83639	Novel human transp
16	19.4	58.8	1977	22	AAH83635	Novel human transp
17	19.4	58.8	2115	22	AAH83636	Novel human transp
18	19.4	58.8	2448	22	AAH83644	Novel human transp
19	19.4	58.8	2586	22	AAH83645	Novel human transp
20	19.4	58.8	2700	22	AAH83640	Novel human transp
21	19.4	58.8	2937	22	AAH83641	Novel human transp
22	19.4	58.8	3075	22	AAH83642	Novel human transp
23	19.4	58.8	3189	22	AAH83637	Novel human transp
24	19.4	58.8	3219	23	ABL21644	Drosophila melanog
25	19.4	58.8	3660	22	AAH83646	Novel human transp
26	19.4	58.8	4149	22	AAH83643	Novel human transp
27	19.4	58.8	4341	22	AAH83647	D. melanogaster la
28	19.4	58.8	5020	22	AAH83647	Novel human transp
29	19.4	58.8	5718	22	AAH83647	CDNA encoding Dros
30	19.4	58.8	5718	23	ABL04853	Drosophila melanog
31	19.4	58.8	9400	23	ABL04852	Drosophila melanog
32	19.2	58.2	380	21	AAH836207	Pinus radiata tran
33	19.2	58.2	393	21	AAH83629	Pinus radiata tran
34	19.2	58.2	5053	21	AAH83626	A. terreus ORF1 en
35	19.2	58.2	16595	22	AAH83641	DNA encoding human
36	19.2	58.2	33000	21	AAH83641	A. terreus DAB seg
37	19.2	58.2	532	22	AAH83649	Human immune/Haema
38	19.2	58.2	532	22	AAH83649	Haemophilus influe
39	18.8	57.0	258	23	AAH83642	Human brain expres
40	18.8	57.0	485	22	AAH83642	Borrelia 2591 OapC
41	18.8	57.0	824	16	AAH83647	Outer surface prot
42	18.8	57.0	1352	22	AAH83647	Genomic sequence #
43	18.8	57.0	3249	22	AAH83647	Human nervous syst
44	18.8	57.0	3249	22	AAH83647	Human reproductive
45	18.8	57.0	3249	22	AAH83647	Human reproductive

ALIGNMENTS

RESULT 1
AAK59770
AAK59770 standard; CDNA; 33 BP.
XX
XX AAK59770;
XX
DT 26-JUL-1999 (first entry)
XX
XX
DE CDNA encoding a thanatine peptide.
XX
KW Thanatine; fungal disease; Cercospora bettula;
KW Cladosporium herbarum; Fusarium curvum; F. graminearum;
KW Phytophthora cinamomi; selection marker; plant transformation;
KW herbicide resistance; ss.
XX
OS Psodius sp.
XX
PN FR2770853-A1.
XX
PD 14-MAY-1999.
XX
PF 07-NOV-1997; 97FR-0014263.
XX
PR 07-NOV-1997; 97FR-0014263.
XX
PA (RHON) RHONE-POULENC AGROCHIMIE.
XX
PI Derose R, Freyssinet G, Hoffmann J;
XX WPI, 1999-315645/27.
DR P-PsDB; AAT13462.
XX

PF New nucleic acid encoding thanatine useful as a selection marker for
PT transformation of plants

XX Claim 3; Page 13; 24pp; French.

CC The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by *Cercospora beticola*, *Cladosporium*
CC *herbarum*, *Fusarium cunorum*, *F. graminearum* and *Phytophthora*
CC *cinamonii*. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the peptide thanatine.

SQ Sequence 33 BP; 10 A; 7 C; 9 G; 7 T; 0 other;

Query Match 100.0%; Score 33; DB 20; Length 33;

Best Local Similarity 100.0%; Pred. No. 0.00014;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcattactctgcacagagagactgtaagtgc 33
DB 1 atcattactctgcacagagagagactgtaagtgc 33

RESULT 2

AA59771 ID AAX59771 standard; cDNA; 63 BP.

XX AAX59771;

DT 26-JUL-1999 (first entry)

XX cDNA encoding a thanatine peptide.

DE Thanatine: fungal disease; bacterial disease; *Cercospora beticola*;

KM *Cladosporium herbarum*; *Fusarium cunorum*; *F. graminearum*;

KW *Phytophthora cinamonii*; selection marker; plant transformation;

XX herbicide resistance; ss.

OS *Psodius* sp.

PN FR2770853-A1.

PD 14-MAY-1999.

PF 07-NOV-1997; 97FR-0014263.

PR 07-NOV-1997; 97FR-0014263.

XX (RHON) RHONE-POULENC AGROCHIMIE.

PI Derose R, Freyssinet G, Hoffmann J;

XX WPI: 1999-315645/27.

DR P-PSDB; AAY15463.

XX New nucleic acid encoding thanatine useful as a selection marker for

PT transformation of plants

PS Claim 4; Page 13; 24pp; French.

XX The specification describes a nucleic acid sequence containing the
XX sequence that encodes thanatine. Plants transformed with this nucleic
XX acid sequence are resistant to fungal and bacterial diseases,
XX specifically those caused by *Cercospora beticola*, *Cladosporium*
XX *herbarum*, *Fusarium cunorum*, *F. graminearum* and *Phytophthora*
XX *cinamonii*. The nucleic acid sequence may also be used as a selection
XX marker for transformation of plants with other coding sequences,
XX e.g. those that impart resistance to herbicides. The present sequence
XX encodes the peptide thanatine.

XX Sequence 63 BP; 19 A; 14 C; 19 G; 11 T; 0 other;

SQ Query Match 100.0%; Score 33; DB 20; Length 63;

Best Local Similarity 100.0%; Pred. No. 0.00016;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcattactctgcacagagagactgtaagtgc 33
DB 22 atcattactctgcacagagagactgtaagtgc 54

RESULT 3

AA59772 ID AAX59772 standard; cDNA; 98 BP.

XX AAX59772;

DT 26-JUL-1999 (first entry)

XX cDNA encoding the signal peptide of the PR-1a gene of tobacco.

DE Thanatine: fungal disease; bacterial disease; *Cercospora beticola*;

KM *Cladosporium herbarum*; *Fusarium cunorum*; *F. graminearum*;

KW *Phytophthora cinamonii*; selection marker; plant transformation;

XX herbicide resistance; PR-1a gene; tobacco; ss.

OS *Nicotiana tabacum*.

PN FR2770853-A1.

PD 14-MAY-1999.

PF 07-NOV-1997; 97FR-0014263.

PR 07-NOV-1997; 97FR-0014263.

XX (RHON) RHONE-POULENC AGROCHIMIE.

PI Derose R, Freyssinet G, Hoffmann J;

XX WPI: 1999-315645/27.

DR P-PSDB; AAY15464.

XX New nucleic acid encoding thanatine useful as a selection marker for

PT transformation of plants

PS Example 1; Page 14; 24pp; French.

XX The specification describes a nucleic acid sequence containing the

XX sequence that encodes thanatine. Plants transformed with this nucleic

XX acid sequence are resistant to fungal and bacterial diseases,

XX specifically those caused by *Cercospora beticola*, *Cladosporium*

XX *herbarum*, *Fusarium cunorum*, *F. graminearum* and *Phytophthora*

XX *cinamonii*. The nucleic acid sequence may also be used as a selection

XX marker for transformation of plants with other coding sequences,

XX e.g. those that impart resistance to herbicides. The present sequence

XX used to make fusion products of the invention.

Query Match 100.0%; Score 33; DB 20; Length 98;

Best Local Similarity 100.0%; Pred. No. 0.00017;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcattactctgcacagagagactgtaagtgc 33
DB 22 atcattactctgcacagagagactgtaagtgc 54

RESULT 4
AAK59774 standard; cDNA; 197 BP.
XX
AC AAK59774;
XX
DT 26-JUL-1999 (first entry)
XX
DE cDNA encoding a thanatine fusion product of the invention.
XX
KW Thanatine; fungal disease; bacterial disease; Cercospora beticola;
KW Cladosporium herbarum; Fusarium culmorum; F. graminearum;
KW Phytophthora cinamomi; selection marker; plant transformation;
KW herbicide resistance; PR-1a gene; tobacco; ss.
XX
OS Synthetic.
XX
FN FR2770853-A1.
XX
PD 14-MAY-1999.
XX
PE 07-NOV-1997; 97FR-0014263.
XX
PR 07-NOV-1997; 97FR-0014263.
XX
PA (RHON) RHONE-POULENC AGROCHIMIE.
XX
PI Derose R, Freyssinet G, Hoffmann J;
XX
XX WPI: 1999-315645/27.
DR P-PSDB; AAY15466.
XX
PT New nucleic acid encoding thanatine useful as a selection marker for
XX transformation of plants
XX
PS Claim 8; Page 14-15; 24pp; French.
XX
XX The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by Cercospora beticola, Cladosporium
CC herbarum, Fusarium culmorum, F. graminearum and Phytophthora
CC cinamomi. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes a thanatine fusion product of the invention, comprising the
CC signal peptide of the PR-1a gene of tobacco.
XX
SQ Sequence 197 BP; 32 A; 56 C; 49 G; 60 T; 0 other;

Query Match 100.0%; Score 33; DB 20; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcatctactgcaacagagagactgtaagtgc 33
|||
Db 123 atcatctactgcaacagagagactgtaagtgc 155

RESULT 5
AAT15600
ID AAT15600 standard; DNA; 3102 BP.
XX
AC AAT15600;
XX
DT 07-APR-1996 (first entry)
XX
DE Laccase-LCC3 gene.
XX
XX Laccase-LCC3; Polyporus pinsitus; Tremetes villosa; primer;
KW polymerase chain reaction; PCR; signal peptide; cellulase;
KW Aspergillus oryzae; cDNA probe; Escherichia coli; plasmid pDSV20;

KW Lignin; lignosulphonate; polymerisation; Kraft pulp;
KW depolymerisation; oxidation; hair dye; phenol; aniline; vector;
KW cloning; Basidiomycetes; ds.
XX
XX Polyporus pinsitus.
OS Tremetes villosa.
XX
FH Key
FT CANT_signal
FT 14..18
FT Location/Qualifiers
FT /*tag= a
FT 406..410
FT TATA_signal
FT /*tag= b
FT 483..2695
FT CDS
FT /*tag= c
FT /product= Laccase-LCC3
FT /note= "EC-1.10.3.2"
FT 483..545
FT sig_peptide
FT /*tag= d
FT 483..665
FT exon
FT /*tag= e
FT 666..720
FT Intron
FT /*tag= f
FT 721..789
FT exon
FT /*tag= g
FT 790..845
FT Intron
FT /*tag= h
FT 846..965
FT exon
FT /*tag= i
FT 966..1020
FT Intron
FT /*tag= j
FT 1021..1134
FT exon
FT /*tag= k
FT 1135..1192
FT Intron
FT /*tag= l
FT 1193..1258
FT exon
FT /*tag= m
FT 1259..1318
FT Intron
FT /*tag= n
FT 1319..1399
FT exon
FT /*tag= o
FT 1400..1460
FT Intron
FT /*tag= p
FT 1461..1616
FT exon
FT /*tag= q
FT 1617..1672
FT Intron
FT /*tag= r
FT 1673..1873
FT exon
FT /*tag= s
FT 1874..1929
FT Intron
FT /*tag= t
FT 1930..1986
FT exon
FT /*tag= u
FT 1987..2036
FT Intron
FT /*tag= v
FT 2037..2237
FT exon
FT /*tag= w
FT 2238..2296
FT Intron
FT /*tag= x
FT 2297..4213
FT exon
FT /*tag= y
FT 2414..2469
FT Intron
FT /*tag= z
FT 2470..2586
FT exon
FT /*tag= {
FT 2587..2638
FT Intron
FT /*tag= |
FT 2639..2695
FT exon
FT /*tag= }
XX
XX WO9600290-A1.
PN
XX
XX 04-JAN-1996.
PD
XX

```

PF 15-JUN-1995; 95MO-US07536.
XX
PR 15-MAY-1995; 95US-0441147.
PR 24-JUN-1994; 94US-0265534.
XX
PA (NOVO ) NOVO NORDISK BIOTECH INC.
XX (NOVO ) NOVO-NORDISK AS.
XX
PI Assaying DA, Dalboge H, Schneider P, Xu F, Yaver DS;
XX WPI; 1996-068874/07.
DR P-PSDB: AAR90723.
XX
PT DNA constructs for expression of Polyopus laccase enzymes - for use
PT in e.g. lignin manipulation, juice mfr., phenol polymerisation and
PT phenol resin prodn
XX
PS Claim 7; Page 67-70; 137pp; English.
XX
CC The sequence encodes laccase-LCC3 (PI 5.23) from Polyopus pinsutus
CC (Trametes villosa). Polymerase chain reaction (PCR) amplification of
CC P. pinsutus cDNA using primers 3331 (AA115603) and 3332 (AA115604) gives
CC a 1500 bp fragment, which is joined to a signal peptide sequence
CC from a 43-kDa cellulase using primer PHD433 (AA115605) and a pUC
CC forward primer in PCR. Clones are expressed in Aspergillus oryzae,
CC and a cDNA probe is obtained and used to screen a P. pinsutus
CC genomic library in Escherichia coli DH5-alpha, giving plasmid
CC pDSY20 (24GEN), with an 8.5-kb HindIII insert (NRRL-B-21267).
CC Screening also results in isolation of LCC1 (AA115598), LCC2
CC (AA115599), LCC4 (AA115601) and LCC5 (AA115602), which encode different
CC laccases produced by P. pinsutus. The laccases may be used to
CC polymerise lignin or lignosulphonates, to depolymerise Kraft pulp,
CC to oxidise dyes or precursors, in hair dye compositions, or to
CC polymerise or oxidise a phenolic or aniline compound. These new
CC laccases are well-expressed in Aspergillus spp. (with vector
CC integration in the genome), in contrast to previous basidiomycete
CC laccases, which give low yields of recombinant enzyme.
XX
SQ Sequence 3102 BP; 669 A; 935 C; 758 G; 739 T; 1 other;

Query Match 67.9%; Score 22.4; DB 17; Length 3102;
Best Local Similarity 81.2%; Pred. No. 9.1;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY . 2 tcactactgcacacagagagactgtaagtgc 33
Db 2623 tcactcttcgacacagagagactgtaacaaagc 2654

RESULT* 6
AAH87754/C
ID AAH87754 standard; cDNA; 589 BP.
XX
XX AAH87754:
XX
AC
XX
XX 25-SEP-2001 (first entry)
XX
DE Peppermint plant oil gland expressed cDNA 110.
XX
XX Peppermint plant oil gland cell; terpenoid essential oil; resin;
XX genetic mapping; antisense suppression; recombinant expression; ss.
XX
XX Mentha x piperita.
XX
XX OS
XX PN WO200153319-A1.
XX
XX 26-JUL-2001.
XX
XX PD
XX PF 19-JAN-2001; 2001MO-US02567.
XX
XX PR 20-JAN-2000; 2000US-0177264.
XX
XX

```

```

PA (CROT/) CROTEAU R B.
PA (LANG/) LANGE B M.
PA (WILD/) WILDUNG M R.
XX
XX Croteau RB, Lange BM, Wildung MR;
XX WPI; 2001-488706/53.
XX
XX
XX New nucleic acid molecules corresponding to mRNA molecules expressed in
XX peppermint oil glands for enhancing expression of plant oil gland cell
XX proteins -
XX
PS Claim 1; Page 119-120; 251pp; English.
XX
XX The invention relates to nucleic acid molecules (AAH87645-AAH88116) that
XX correspond to all or part of a mRNA molecule expressed in plant oil
XX gland cells, especially peppermint and plant oil glands that produce
XX terpenoid essential oils and resins. The nucleic acids are useful for
XX genetically mapping a plant genome for genes expressed in plant oil
XX gland cells and to suppress (for example by antisense suppression) or
XX enhance their expression (for example by genetically transforming a
XX plant cell with a replicable expression vector that expresses one or more
XX proteins naturally expressed in plant oil gland cells). The nucleic acids
XX are also useful for recombinant expression of plant oil gland proteins
XX required for terpenoid essential oil and/or resin production in bacterial
XX and/or yeast cells.
XX
SQ Sequence 589 BP; 164 A; 115 C; 135 G; 175 T; 0 other;

Query Match 61.2%; Score 20.2; DB 22; Length 589;
Best Local Similarity 75.8%; Pred. No. 56;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY . 1 atcatctacgcacacagagagactgtaagtgc 33
Db 396 ATCATCTCCACAGGAACACACACACTGTAACACG 364

RESULT 7
AAV83948/C
ID AAV83948 standard; DNA; 18073 BP.
XX
XX AAV83948:
XX
AC
XX
XX 03-MAR-1999 (first entry)
XX
DE Bacterial artificial chromosome (BAC)-F2 contig 8.
XX
XX Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
XX neocentromere; replication; extra-chromosomal element; segregation;
XX cell division; artificial chromosome; gene therapy; BAC; transgenic;
XX human artificial chromosome; bacterial artificial chromosome; ss.
XX
XX Synthetic.
XX
XX OS
XX PN WO9851790-A1.
XX
XX 19-NOV-1998.
XX
XX PF 13-MAY-1998; 98MO-AU00352.
XX
XX PR 26-AUG-1997; 97AU-0008791.
XX
XX PR 13-MAY-1997; 97AU-0006784.
XX
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
XX Cancellia MR, Choo K, Du Sart D;
XX
XX WPI; 1999-009773/01.
XX
XX 'New isolated nucleic acid comprising neocentromere sequences from
XX eukaryotic chromosome - used to produce replicable, segregating
XX

```


PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy
XX
PS Claim 10; Page 209-222; 540pp; English.
XX
CC The present sequence represents a bacterial artificial chromosome (BAC)
CC contig, and exemplifies the invention. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in
CC a compatible cell, of replicating, acting as extra-chromosomal element
CC and segregating during cell division. The sequences can be used to
CC construct artificial chromosomes for use in gene therapy comprising a
CC replicable, segregating nucleic acid that confers a specific phenotype
CC on cells. Human artificial chromosomes can propagate in human cells and
CC carry large amounts of DNA (e.g. therapeutic genes) and, being
CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC are also useful for generation of transgenic plants and animals, in
CC production of proteins and to make diagnostic reagents, e.g. for
CC expression of cytokines, receptors and growth factors, or to increase
CC the copy number of a gene in a cell. The constructs may also be
CC used for functional and structural analysis of chromosomes.
XX
SQ Sequence 18073 BP; 4755 A; 3584 C; 3589 G; 6145 T; 0 other:

Query Match 61.2%; Score 20.2; DB 20; Length 18073;
Best Local Similarity 88.0%; Pred. No. 1e+02; Mismatches 0; Gaps 0;
Matches 22; Conservative 0; Indels 3; Indels 0; Gaps 0;

QY 8 actgcacacagagagactgtaagt 32
||| ||||| ||||| ||||| ||
Db 8535 ACTGTTACAGCAGAGACTGTAATG 8511

RESULT 8
AAS82520 standard; cDNA; 1275 BP.
ID AAS82520
AC AAS82520;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #18324.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB: ABG18333.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 18324; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1275 BP; 293 A; 328 C; 361 G; 293 T; 0 other:

Query Match 60.6%; Score 20; DB 23; Length 1275;
Best Local Similarity 82.1%; Pred. No. 79;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 attactgcacacagagagactgtaagt 31
||||| ||||| ||| || ||||| ||
Db 205 attacttcaactcgtgaagagcgtgaagt 232

RESULT 9
AAS82525 standard; cDNA; 1737 BP.
ID AAS82525
AC AAS82525;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #18329.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB: ABG18338.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 18329; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX

SO Sequence 1737 BP; 449 A; 405 C; 462 G; 421 T; 0 other;

Query Match 60.6%; Score 20; DB 23; Length 1737;
Best Local Similarity 82.1%; Pred. No. 83;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 atctactgcacagagagactgtaagt 31
||||| ||||| ||| ||| |||||
Db 988 atctactgcacagagagactgtaagt 1015

RESULT 10

AAS89143
ID AAS89143 standard; CDNA: 2289 BP.

XX AAS89143:

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #24947.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG24956.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits and to assess

XX biodiversity -

XX Claim 1; SEQ ID No 24947; 103pp; English.

CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX

SO Sequence 2289 BP; 516 A; 595 C; 678 G; 500 T; 0 other;

Query Match 60.6%; Score 20; DB 23; Length 2289;
Best Local Similarity 82.1%; Pred. No. 87;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 atctactgcacagagagactgtaagt 31
||||| ||||| ||| ||| |||||
Db 493 atctactgcacagagagactgtaagt 520

RESULT 11

AAS48172
ID AAS48172 standard; DNA: 179 BP.

XX AAS48172:

XX 13-FEB-2002 (first entry)

XX Enterococcus faecalis cellular proliferation inhibitory sequence #742.

XX Antisense; ss; prokaryotic cellular proliferation;

XX antibiotic; antibacterial; drug design.

XX Enterococcus faecalis.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX WPI; 2001-611495/70.

XX New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 1; Seq ID No 749; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence is an antisense
CC oligonucleotide of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 179 BP; 54 A; 38 C; 30 G; 57 T; 0 other;

Query Match 60.0%; Score 19.8; DB 23; Length 179;
Best Local Similarity 77.4%; Pred. No. 67;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 atcattactgcacagagagactgtaagt 31
|||||
DB 43 atcattactgcacagagagactgtaagt 73

RESULT 12
AAC26094/C
ID AAC26094 standard; CDNA; 404 BP.
XX
AC AAC26094;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 30169.

XX Human; 5' EST; expressed sequence tag; secreted protein; CDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX

OS Homo sapiens.

PN EPI033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 30169; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences

CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
XX
SQ Sequence 404 BP; 132 A; 62 C; 72 G; 136 T; 2 other;

Query Match 59.4%; Score 19.6; DB 21; Length 404;
Best Local Similarity 84.6%; Pred. No. 94;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 catctactgcacagagagactgta 28
|||||
DB 39 CATCTGCTGCACAGAGATGCTGTA 14

RESULT 13
ABL21645/C
ID ABL21645 standard; DNA; 792 BP.
XX
AC ABL21645;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 16408.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-655686/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX

PS Claim 1; SEQ ID NO 16408; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABBS7737-ABBS72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 792 BP; 200 A; 258 C; 190 G; 144 T; 0 other;

Query Match 58.8%; Score 19.4; DB 23; Length 792;
Best Local Similarity 79.3%; Pred. No. 1.3e+02;

```
Matches 23: Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 tcatctactgcacagagagactgtaag 30
   ||||| ||||| ||||| ||||| |||||
DB 353 TCATCCACTGCACGGGAGAGATGAGAAG 325

RESULT 14
AAF83638
ID AAF83638 standard; CDNA; 1488 BP.
XX
AC AAF83638;
XX
DT 23-JUL-2001 (first entry)
XX
DE Novel human transporter protein (NHP) encoding CDNA.
XX
KM Novel human protein; transporter protein; NHP; therapeutic; diagnostic;
XX gene therapy; ss.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT CDS 1..1488
FT /*tag= a
XX
PN MO200132706-A2.
XX
PD 10-MAY-2001.
XX
PF 31-OCT-2000; 2000MO-US29852.
XX
PR 02-NOV-1999; 99US-0163018.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Harris M, Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
DR MPI; 2001-343477/36.
DR P-PSDB; AAB62550.
XX
XX
PT Novel isolated human polynucleotide sequences encoding polypeptides
PT that share sequence similarity with mammalian multidrug resistance
PT proteins and cellular transporter proteins, useful as probe or primer
PT ..
XX
PS Disclosure; Page 32-33; 59pp; English.
XX
CC The invention relates to novel human transporter proteins (NHP) and
CC polynucleotides encoding the NHPs. Fragments of NHP nucleotide sequences
CC can be used in conjunction with PCR to screen libraries, isolate clones
CC and prepare cloning and sequencing templates. The NHP oligonucleotides
CC can also be used as hybridization probes for screening libraries, for
CC assessing gene patterns and for preparing antisense nucleic acid
CC molecules. The NHP nucleotide sequences are also useful in screening
CC techniques for drugs which treat symptomatic or phenotypic
CC manifestations of perturbing the normal function of NHP in the body.
CC Sequences AAF83635-647 represent NHP nucleic acid sequences.
XX
SQ Sequence 1488 BP; 361 A; 367 C; 414 G; 346 T; 0 other;

Query Match 58.88; Score 19.4; DB 22; Length 1488;
Best Local Similarity 79.3%; Pred. NO. 1.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 5 tctactgcacagagagactgtaagtc 33
   ||| | ||||| | ||| ||||| ||
DB 668 tctgcgcaacacgcgggagtgttaagagc 696

RESULT 15
```

```
AAF83639
ID AAF83639 standard; CDNA; 1626 BP.
XX
AC AAF83639;
XX
DT 23-JUL-2001 (first entry)
XX
DE Novel human transporter protein (NHP) encoding CDNA.
XX
KM Novel human protein; transporter protein; NHP; therapeutic; diagnostic;
XX gene therapy; ss.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT CDS 1..1626
FT /*tag= a
XX
PN MO200132706-A2.
XX
PD 10-MAY-2001.
XX
PF 31-OCT-2000; 2000MO-US29852.
XX
PR 02-NOV-1999; 99US-0163018.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Harris M, Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
DR MPI; 2001-343477/36.
DR P-PSDB; AAB62551.
XX
XX
PT Novel isolated human polynucleotide sequences encoding polypeptides
PT that share sequence similarity with mammalian multidrug resistance
PT proteins and cellular transporter proteins, useful as probe or primer
PT -
XX
PS Disclosure; Page 34-36; 59pp; English.
XX
CC The invention relates to novel human transporter proteins (NHP) and
CC polynucleotides encoding the NHPs. Fragments of NHP nucleotide sequences
CC can be used in conjunction with PCR to screen libraries, isolate clones
CC and prepare cloning and sequencing templates. The NHP oligonucleotides
CC can also be used as hybridization probes for screening libraries, for
CC assessing gene patterns and for preparing antisense nucleic acid
CC molecules. The NHP nucleotide sequences are also useful in screening
CC techniques for drugs which treat symptomatic or phenotypic
CC manifestations of perturbing the normal function of NHP in the body.
CC Sequences AAF83635-647 represent NHP nucleic acid sequences.
XX
SQ Sequence 1626 BP; 398 A; 404 C; 442 G; 382 T; 0 other;

Query Match 58.88; Score 19.4; DB 22; Length 1626;
Best Local Similarity 79.3%; Pred. NO. 1.5e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 5 tctactgcacagagagactgtaagtc 33
   ||| | ||||| | ||| ||||| ||
DB 668 tctgcgcaacacgcgggagtgttaagagc 696

Search completed: May 12, 2002, 20:20:36
Job time: 4923 sec
```


BASE COUNT 10 a 7 c 9 g 7 t

ORIGIN

Query Match 100.0%; Score 33; DB 6; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcactctactgcacagggagactgtaagtgc 33
|||||
Db 1 ATCATCTACTGCACAGGAGGACTGTAAGTGC 33

RESULT 2

A95924 63 bp DNA linear PAT 26-JAN-2000
LOCUS A95924
DEFINITION Sequence 2 from Patent WO924594.
ACCESSION A95924
VERSION A95924.1 GI:6779828
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
1 (bases 1 to 63)
REFERENCE
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 924594-A 2 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source
1. 63
/organism="unidentified"
/db_xref="taxon:32644"
<1..>63
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70492.1"
/db_xref="GI:6779829"
/translation="GSKRPVPIIYCNRRGKCQRM"

BASE COUNT 19 a 14 c 19 g 11 t

ORIGIN

Query Match 100.0%; Score 33; DB 6; Length 63;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcactctactgcacagggagactgtaagtgc 33
|||||
Db 22 ATCATCTACTGCACAGGAGGACTGTAAGTGC 54

RESULT 3

A95931 87 bp DNA linear PAT 26-JAN-2000
LOCUS A95931
DEFINITION Sequence 9 from Patent WO924594.
ACCESSION A95931
VERSION A95931.1 GI:6779839
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
1 (bases 1 to 87)
REFERENCE
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 924594-A 9 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source
1. 87
/organism="unidentified"

BASE COUNT 14 a 31 c 21 g 21 t

ORIGIN

Query Match 100.0%; Score 33; DB 6; Length 87;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcactctactgcacagggagactgtaagtgc 33
|||||
Db 77 ATCATCTACTGCACAGGAGGACTGTAAGTGC 45

RESULT 4

A95925 98 bp DNA linear PAT 26-JAN-2000
LOCUS A95925
DEFINITION Sequence 3 from Patent WO924594.
ACCESSION A95925
VERSION A95925.1 GI:6779830
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
1 (bases 1 to 98)
REFERENCE
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 924594-A 3 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source
1. 98
/organism="unidentified"
/db_xref="taxon:32644"
<1..>66
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70493.1"
/db_xref="GI:6779831"
/translation="GSKRPVPIIYCNRRGKCQRM"

BASE COUNT 25 a 23 c 34 g 16 t

ORIGIN

Query Match 100.0%; Score 33; DB 6; Length 98;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcactctactgcacagggagactgtaagtgc 33
|||||
Db 22 ATCATCTACTGCACAGGAGGACTGTAAGTGC 54

RESULT 5

A95927 197 bp DNA linear PAT 26-JAN-2000
LOCUS A95927
DEFINITION Sequence 5 from Patent WO924594.
ACCESSION A95927
VERSION A95927.1 GI:6779834
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
1 (bases 1 to 197)
REFERENCE
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 924594-A 5 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source
1. 197
/organism="unidentified"

CDS
/db_xref="taxon:32644"
12.167
/note="unnamed protein product"
/codon_start=1
/protein_id="CA870495.1"
/db_xref="GI:6779835"
/translation="MGFLFSQLPSLLVSTLLFLVISHSCRAGSKKVPYIYNRR
TGCGQRN"

BASE COUNT 32 a 56 c 49 g 60 t
ORIGIN

Query Match 100.0%; Score 33; DB 6; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atcactctacgcaagagagactgtaagtc 33
Db 123 ATCATCTACTGCAACAGAGAGACTGTAAGTC 155

RESULT 6
AF170051 540 bp DNA linear VRL 05-MAR-2001
LOCUS HIV-2 isolate 96207 from France envelope glycoprotein (env) gene,
DEFINITION partial cds.
ACCESSION AF170051
VERSION AF170051.1 GI:11990555
KEYWORDS Human immunodeficiency virus type 2.
SOURCE Human immunodeficiency virus type 2.
ORGANISM HIV-2 isolate 96207 from France
1 (bases 1 to 540)
Virus; Retrovirus; Retroviridae; Lentivirus; Primate
lentivirus group.

REFERENCE
AUTHORS Diamond, P., Apelet, C., Robertson, D. L., Souquiere, S., Lepretre, A.,
TITLE Matchon, S., Plantier, J. C., Brun-Vezinet, F., and Simon, F.
JOURNAL Infecting patients living in France
PUBMED 1162815
Virology 280 (1), 19-30 (2001)
2 (bases 1 to 540)
AUTHORS Francois, S., Diamond, P., Apelet, C., Robertson, D. L., Souquiere, S.,
TITLE Lepretre, A., Matchon, S., and Brun-Vezinet, F.
JOURNAL Direct Submission
PUBMED 1162815
Submitted (19-JUL-1999) Laboratoire de Virologie et Services des
Maladies Infectieuses et Tropicales, Hôpital Bichat - Claude
Bernard, Paris 75877 Paris Cedex 18, France
location/Qualifiers

FEATURES
Source
1.540
/organism="Human immunodeficiency virus type 2"
/isolate="96207"
/db_xref="taxon:11709"
/country="France"
/gene="env"
/gene="env"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAG42191.1"
/db_xref="GI:11990555"
/translation="DWSTFETSIKPCVKILPLICVAMNCTKPSNTTATSTST
STATGAVNENPCVSTNCTGEEEMITCNFSMGLODKKRRNENYKSDYIC
ESNNIDGNISRCYMSHCNTVTITSCDKNYDTRFRCAPGVALRCNDYNTG
FAVNCNKVVAATCTRMETQ"

BASE COUNT 190 a 115 c 110 g 125 t
ORIGIN

Query Match 67.9%; Score 22.4; DB 14; Length 540;
Best Local Similarity 81.2%; Pred. No. 16;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 atcactctacgcaagagagactgtaagtc 32
Db 132 ATCAACACTGCACTGAGAGAGTGAATG 163

RESULT 7
165231 3102 bp DNA linear PAT 07-OCT-1997
LOCUS Sequence 5 from patent US 5667531.
DEFINITION 165231
ACCESSION 165231
VERSION 165231.1 GI:2461801
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 3102)
AUTHORS Yaver, D. Sue., Xu, F., Dalb.o slashed, ge, H., Schneider, P. and
TITLE Aaslyng, D. A.
JOURNAL Dye compositions containing purified polyporus laccases and nucleic
FEATURES acids encoding same
SOURCE Patent: US 5667531-A 5 16-SEP-1997;
Location/Qualifiers
1.3102
/organism="unknown"

BASE COUNT 669 a 935 c 758 g 739 t 1 others
ORIGIN

Query Match 67.9%; Score 22.4; DB 6; Length 3102;
Best Local Similarity 81.2%; Pred. No. 19;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 tcatctactgcaagagagactgtaagtc 33
Db 2623 TCATCTTTTGCAACAGAGAGACTGCAACAGC 2654

RESULT 8
AC095159 157853 bp DNA linear HTG 20-DEC-2001
LOCUS Rattus norvegicus clone CH230-8C21, *** SEQUENCING IN PROGRESS ***
DEFINITION 62 unordered pieces.
ACCESSION AC095159
VERSION AC095159.2 GI:17942168
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 157853)
AUTHORS Muzny, D. M., Adams, C., Adio-Oduola, B., Ali-osman, F. R., Allen, C.,
Alsbrooks, S. L., Amaralunga, H. C., Are, J. R., Banks, T., Barbara, J.,
Benton, J., Bimaga, K., Blankenburg, K., Bonin, D., Bouch, J.,
Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N. P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C., Carron, T. F.,
Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C. D., Cox, C.,
Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, O.,
Day-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O.,
Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H.,
Duan, R., Rocha, S., Durbin, K. J., Earnhart, C., Edgar, D., Edwards, C. C.,
Elhaj, C., Escotto, M., Falls, T., Ferriguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Gaza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C.,
Hollins, B., Homs, J., Howard, S., Huber, J., Huyl, S., Hume, J.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulseged, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R.,

Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawliny, E., McLeod, M. P., Meador, M.,
Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabhat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S.,
Oguni, M., Okunolu, G., Oragunye, N., Oviator, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Plickens, R., Primus, E., Fu, L. L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojuben, I., Rolfe, M.,
Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, P., Telford, B., Thomas, N.,
Thomas, S., Uemari, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Washington, S., Williams, G., Williamson, A., Wlarczyk, R., Woodson, S.,
Wortley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
2 (bases 1 to 157853)
Unpublished

Wortley, K. C.
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced g1:15625713.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GCVH
Center clone name: CH230-8C21
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329First call to
findhaplist
Consensus quality: 124710 bases at least Q40
Consensus quality: 132941 bases at least Q30
Consensus quality: 139599 bases at least Q20
Estimated insert size: 121782; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 10809: contig of 10809 bp in length
* 10810 10909: gap of unknown length
* 10910 16007: contig of 5098 bp in length
* 16008 16107: gap of unknown length
* 16108 20377: contig of 4270 bp in length
* 20378 20477: gap of unknown length
* 20478 25692: contig of 5215 bp in length
* 25693 31341: contig of 5549 bp in length
* 31342 31441: gap of unknown length
* 31442 38548: contig of 7107 bp in length
* 38549 38648: gap of unknown length
* 38649 42503: contig of 3855 bp in length
* 42504 42603: gap of unknown length
* 42604 46221: contig of 3618 bp in length
* 46222 46321: gap of unknown length
* 46322 50780: contig of 4459 bp in length
* 50781 50880: gap of unknown length
* 50881 54670: contig of 3790 bp in length

54671 54770: gap of unknown length
* 54771 57876: contig of 3106 bp in length
* 57877 57976: gap of unknown length
* 57977 61380: contig of 3404 bp in length
* 61381 61480: gap of unknown length
* 61481 63945: contig of 2465 bp in length
* 63946 64045: gap of unknown length
* 64046 65871: contig of 1826 bp in length
* 65872 65972: gap of unknown length
* 65973 69367: contig of 3336 bp in length
* 69368 72572: contig of 3105 bp in length
* 72573 72673: gap of unknown length
* 72674 75793: contig of 3121 bp in length
* 75794 75893: gap of unknown length
* 75894 77882: contig of 1989 bp in length
* 77883 79716: contig of 1734 bp in length
* 79717 79817: gap of unknown length
* 79818 82887: contig of 3071 bp in length
* 82888 82987: gap of unknown length
* 82988 85212: contig of 2225 bp in length
* 85213 87069: contig of 1757 bp in length
* 85313 87169: gap of unknown length
* 87070 89966: contig of 2797 bp in length
* 89967 90067: gap of unknown length
* 90068 92128: contig of 2062 bp in length
* 92129 92229: gap of unknown length
* 92230 94503: contig of 2275 bp in length
* 94504 94603: gap of unknown length
* 94604 97620: contig of 3017 bp in length
* 97621 97721: gap of unknown length
* 97722 100438: contig of 2718 bp in length
* 100439 100538: gap of unknown length
* 100539 102722: contig of 2184 bp in length
* 102723 102822: gap of unknown length
* 102823 104843: contig of 2021 bp in length
* 104844 104943: gap of unknown length
* 104944 107122: contig of 2179 bp in length
* 107123 109547: contig of 2335 bp in length
* 109548 109647: gap of unknown length
* 109648 111155: contig of 1508 bp in length
* 111156 111255: gap of unknown length
* 111256 113579: contig of 2324 bp in length
* 113580 113679: gap of unknown length
* 113680 115794: contig of 2115 bp in length
* 115795 115894: gap of unknown length
* 115895 117972: contig of 2078 bp in length
* 117973 118072: gap of unknown length
* 118073 119744: contig of 1672 bp in length
* 119745 119844: gap of unknown length
* 119845 121302: contig of 1458 bp in length
* 121303 121402: gap of unknown length
* 121403 122766: contig of 1364 bp in length
* 122767 122866: gap of unknown length
* 122867 124357: contig of 1491 bp in length
* 124358 124457: gap of unknown length
* 124458 125917: contig of 1460 bp in length
* 125918 126017: gap of unknown length
* 126018 127778: contig of 1761 bp in length
* 127779 127878: gap of unknown length
* 127879 129566: contig of 1688 bp in length
* 129567 129666: gap of unknown length
* 129667 131623: contig of 1957 bp in length
* 131624 131723: gap of unknown length
* 131724 132811: contig of 1188 bp in length
* 132812 133011: gap of unknown length
* 133012 134523: contig of 1512 bp in length
* 134524 134623: gap of unknown length
* 134624 136223: contig of 1600 bp in length
* 136224 136323: gap of unknown length

*	136324	137533	contig of 1210 bp in length
*	137534	137633	gap of unknown length
*	137634	138659	contig of 1026 bp in length
*	138660	138759	gap of unknown length
*	138760	139983	contig of 1230 bp in length
*	139990	140089	gap of unknown length
*	140090	141463	contig of 1374 bp in length
*	141464	141563	gap of unknown length
*	141564	143038	contig of 1475 bp in length
*	143039	143138	gap of unknown length
*	143139	144159	contig of 1021 bp in length
*	144160	144259	gap of unknown length
*	144260	145363	contig of 1104 bp in length
*	145364	145463	gap of unknown length
*	145464	146750	contig of 1287 bp in length
*	146751	146850	gap of unknown length
*	146851	147939	contig of 1149 bp in length
*	148000	148099	gap of unknown length
*	148100	149878	contig of 1779 bp in length
*	149879	149978	gap of unknown length

	Query Match	66.7%	Score	22	DB	2	Length	15785;
b	Best Local Similarity	83.3%	Pred.	No	44			
b	Matches	25	Conservative	0	Mismatches	5	Indels	0; Gaps
QY	1 atcattactgcaacaggaggaactgttaag	30						
b	27901 ACCGATATAGCTCAACAGCAGAACACTCGGAAG	27930						

[illegible]

/protein_id="AAC98354.1"
/db_xref="GI:4063639"
/translation="MMNCIRIRNMPLPACLLITIMTKYSSSGFIELDLAVENVNGI
LMDECDSDSTNSQDQRCVRDECDTVEYVCLKEQSEYTTTGGCTFFSGSDYDLGNI
ISFTAKNSPKSTEDVGRKIIPFHPAMRSYTLILEADMMSTNNQNEENIIEHII
ASMYNPGHMMSEIRHPGITAHIEYRIPRCENYVSGSKCNQCRADRDYEGYRCDP
GNIYCLDGMMGDEDRATACROGCNLIHGGCAVPKCGKCNMOGQFCEDELPCCL
GTCMPMPNCTCEKRWGGILCDKDLNPGCTHHPCVNGTGCMSNPENYACAPCEYSGG
NCEIAEHACVSNPACNAGTGTCEVPLGFEGCHCPKQHEGPTCAKDMECCASPRAGSGT
IDLEHGEFCVCPQMVNGTKTCQIDANECKMGCPVNAHSGNNIGTGHCCDFQMAQGN

BASE COUNT	1296 a	1224 c	1360 g	1258 t
ORIGIN				
Query Match		64.8%	Score 21.4	DB 5; Length 5138;
Best Local Similarity		80.6%	Pred. No. 58;	
Matches 25; Conservative		0;	Mismatches 6;	Indels 0; Gaps 0;
QY	2	tcac	tctactgcacacagagagactggtgaatg	32
db	4908	TTATGATCAGCAACAGGAGGACTGTTAAGGG	4878	

RESULT	10
AF229450/c	
LOCUS	
DEFINITION	Danio rerio jagged2 (jag2) mRNA, complete cds, alternatively spliced.
ACCESSION	AF229450
VERSION	AF229450.1
KEYWORDS	GI:15799277
SOURCE	.
ORGANISM	zebrafish. Danio rerio Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio. 1 (bases 1 to 5355) Oda,T., and Chandrasekharappa,S.C. Isolation, Characterization and Expression Analysis of Zebrafish Jagged Genes Unpublished 2 (bases 1 to 5355) Oda,T., and Chandrasekharappa,S.C. Direct Submission Submitted (31-JAN-2000) GMB/MHGRT, NIH, 49 Convent Drive, Bethesda, MD 20852-4442, USA Location/Qualifiers 1..5355
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
FEATURES	
SOURCE	

FEATURES	source	Location/Qualifiers
		1..5355
		/organism="Danio rerio"
		/db_xref="taxon:7955"
		/chromosome="LGX"
gene		1..5355
		/gene="jag2"
		148..3798
		/gene="jag2"
CDS		/note="Notch ligand; alternatively spliced"
		/codon_start=1
		/product="jagged2"
		/protein_id="AI08215.1"
		/db_xref="GI:15799278"
		/translation="MNCRCVRINMLPFCILLLIMTKYSQSGFYELQIAVENNGI LMDGCCDSTRNSQQRVROECDYFVCLKEQSEVYTGQCFGSGSDYLDGN PSFKPAKNSPSKTSVGRKILIPHPAMPBSYLLILEAWDMDSITNNSEENILEHRI ASMPAGDHMQSIRHPGTAHLEVIRVRCDENYGSKCNKQCRPRDYFGYRDPD GNIVCLDGMGDCRTAICQKQCNLIHGCAVPGECCKNYGQSFCECLYFPCCL GTGVCQCTCEKNNMGLCDKDLNYCGTHHCVGNGTCMNEPEPYACAPGEGSGI NCEIAEHGCVSNPCANGGTEHPGVEGHCPCADMDQCAKDECSAPCAOGCTG IDLNEFFCVCPQMYVCKTQDIDLNGCHGCGCONAKTEIKHGVGHCCQAFVGLH EVRNRKCSGCPQNGRCHVILDSFYVCECPSPSYACMLEVESLHPNRCERPQONT

Query Match	64.8%	Score 21.4	DB 5	Length 5355
Best Local Similarity	80.6%	Pred. No. 58		
Matches	25	Conservative	0	Mismatches 6; Indels 0; Gaps 0;
2	tcatctactgcacagagagagactgtgtagtg	32		
1				
Db	4816	TTATGATCAGCAGACAGAGAGACTGCTAAGG	4786	
RESULT 11	AF229449	5469 bp	mRNA	linear
LOCUS	AF229449/c			VRT 28-SEP-2001
DEFINITION	Danio rerio jagged2 (jag2) mRNA, complete cds, alternatively			
ACCESSION	AF229449			
VERSION	AF229449.1	GI:15799275		
KEYWORDS				
SOURCE	zebrafish.			
ORGANISM	Danio rerio			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.			
AUTHORS	Oda, T. and Chandrasekharappa, S.C.			
REFERENCE	Isolation, Characterization and Expression Analysis of Zebrafish Jagged Genes			
TITLE	Unpublished			
JOURNAL	2 (bases 1 to 5469)			
REFERENCE	Oda, T. and Chandrasekharappa, S.C.			
AUTHORS	Direct Submission			
TITLE	Submitted (31-JAN-2000) GMBB/NHGRI, NIH, 49 Convent Drive, Bethesda, MD 20892-4442, USA			
JOURNAL	location/Qualifiers			
FEATURES	1..5469			
source	/organism="Danio rerio"			
gene	/db_xref="taxon:7955"			
CDs	/chromosome="LGX"			
	1..5469			
	/gene="jag2"			
	148..3912			
	/gene="jag2"			
	/note="Notch ligand; alternatively spliced"			
	/codon_start=1			
	/product="jagged2"			
	/protein_id="AA108214.1"			
	/db_xref="GI:15799276"			
	/translation="MNMCIIRMLPLPCLLLTMMTVSSSGVFELQILAVENGEELMGECDSTRNSODRCVRCDEQTPYKCLLEKQVSTGTTGCTPPSSSGSGVGLMIFSRTRAKNSPKSDVCKITIPFRFAMPRTSTLTLEAMDNSTONNGEENLIRHITHASNVPGDHWOSIRHPGITAHIERIVARCDENYGSCKMCKRPDRDYFGHYCDPSGNTVCLDGMGDEIRCPILGICQGNMLIHGGCAVPDECKCNGMOGQFDECLPYGCLHAGTCVMPQCTCEKMGGLLDDKDLNYCGTHHPCVNGGTCMNSBEDENACAPDEYSGKNCETLAHAGVSNPCANGGTGHEVPTGEGCHPCPMEDPTCAKMDDEACSSPCAGGTCINLENGECVCPQMWGKTGQIOIANECMGKPCVAHASKNIGGYHDCFGMAQMCIDINLEGGHCGQCNAGATKRELYHGTYHCCCPAGFVGLCEVSRNKCASGCPQANGRCYVILDSFVCECSNTAGMLCEVESUSHNPCEPNCQNTALCTSLPGDFTYCAPEYBCK			

polyA_signal

```
4583..4588
/gene=""jaag2"
/note="alternate; results in shorter message"
```

polyA_signal

```
5399..5404
/gene=""jaag2"
```

BASE COUNT	1393 a	1264 c	1430 g	1382 t
ORIGIN				

Query Match 64.8% Score 21.4; DB 5; Length 5469;
Best Local Similarity 80.6%; Pred No. 58;

Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Dy 2 tcattctactgcacaggaggaagtgttaagtg 32
 || | | | | | | | | | | | | | |
Db 4930 TTATGATCAGCAACGACGAGACTGGTAGG 4900

RESULT 12

ACC005792 LOCUS 81357 bp DNA linear PRI 06-OCT-1998

DEFINITION Homo sapiens chromosome 19, BAC CIT-B-978SK-259j21, complete sequence.

ACCESSION ACC005792

VERSION ACC005792.1 GI:3702276

KEYWORDS HNG.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 81357)
Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Viswanathan,V., Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stillwagen,S., Phan,H., Velasco,N., Do,L., Regala,R., Terry,A., Barnes,J., Dangnan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J., Liu,S., Altix,C., Andreise,T., Frankheim,M., Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Krommiller,B., Arellano,A., Saunders,C., Ow,D., Nolan,M., Trong,S., Komiyashi,A., Olsen,A.S. and Carraro,A.V.

AUTHORS Sequence analysis of a 600 bp human contig in 19q13.1 between D19S213 and D19S9

TITLE Unpublished

JOURNAL 2 (bases 1 to 81357)

REFERENCE Lamerdin,J.E.

AUTHORS Direct Submission

JOURNALS Submitted (06-Oct-1998) Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA

COMMENT Map and sequence oriented from q centimere to telomere. BAC 249895 (CIT-B-978SK-259j21) is expected to overlap BAC 256788 (CIT-B-978SK-277120) to the left by approximately 12 kb. BAC 249895 is currently the telomeric end of this sequencing tilting path. Additional chromosome 19 map and sequence information are available from: <http://www.bio.lnl.gov/genome/genome.html>.

FEATURES
Source Location/Oualifiers

1..81357 /organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.1 between D19S213 and D19S9"
/cclone="BAC CIT-B-978SK-259j21 (BC249895)"
/cell_line="978SK"

```
repeat_region /clone.lib="Cal tech CIT-B BAC library"
              /rpt_family="AluJo"
              complement(1824..1846)
repeat_region /rpt_family="AT-rich"
              complement(1912..1997)
repeat_region /rpt_family="7SLRNA"
              complement(2009..2309)
repeat_region /rpt_family="AluY"
              complement(2312..2531)
              /rpt_family="7SLRNA"
              complement(3052..3144)
misc_feature /note="predicted exon, program: gra112exons_human_1.3,
             frame: 0, quality: good, score: 50.000"
repeat_region /rpt_family="AluJo"
              complement(4576..4873)
repeat_region /rpt_family="AluJo"
              complement(4888..5416)
              /rpt_family="LINE2"
              complement(5428..5654)
              /rpt_family="AluJo"
              complement(5655..5956)
repeat_region /rpt_family="AluY"
              complement(5959..6039)
              /rpt_family="AluJo"
              complement(6571..6888)
              /rpt_family="AluSx"
              complement(7532..7568)
              /rpt_family="(TAAAn)"
              complement(7569..7872)
              /rpt_family="AluSx"
              complement(8178..8220)
              /rpt_family="(CA)n"
              complement(8221..8510)
              /rpt_family="AluJb"
              8741..8945
misc_feature /note="predicted exon, program: gra112exons_human_1.3,
             frame: 2, quality: good, score: 52.000"
repeat_region /rpt_family="MIR"
              complement(11130..11264)
              /rpt_family="FLAM_C"
              12397..13128
              /rpt_family="L1PB1"
              complement(13140..13162)
              /rpt_family="AT-rich"
              complement(13323..13617)
              /rpt_family="L1MB4"
              complement(13670..13715)
              /rpt_family="MIR"
              complement(14717..15020)
              /rpt_family="AluJo"
              complement(15207..15253)
              /rpt_family="Alu"
              15264..15332
              /rpt_family="LINE2"
              complement(15352..15649)
              /rpt_family="AluY"
              complement(15740..16136)
              /rpt_family="MSRA"
              complement(16332..16629)
              /rpt_family="AluJo"
              complement(16934..17299)
              /rpt_family="MLT1D"
              complement(17315..17577)
              /rpt_family="AluY"
              complement(17589..17886)
              /rpt_family="AluSx"
              complement(17896..17974)
              /rpt_family="MLT1D"
              complement(18949..19125)
              /note="BLASTN similarity to 222632 (87..263); match: 1,
              score: 4.0e-65; database searched: nt; H.sapiens repeat
              region DNA."
```

```
repeat_region complement(19132..19428)
              /rpt_family="AluSx"
              complement(20228..20268)
              /rpt_family="(TAAAn)"
              complement(20269..20570)
              /rpt_family="AluSx"
              20898..21065
              /rpt_family="MSTB"
              21066..21295
              /rpt_family="AluSx"
              21313..21479
              /rpt_family="MSTB"
              complement(21480..21785)
              /rpt_family="AluJo"
              21787..21851
              /rpt_family="MSTB"
              complement(23323..23618)
              /rpt_family="AluSx"
              24242..24346
misc_feature /note="predicted exon, program: gra112exons_human_1.3,
             frame: 0, quality: good, score: 73.000"
repeat_region /rpt_family="AT-rich"
              complement(24579..24599)
              /rpt_family="AT-rich"
              complement(24877..24915)
              /rpt_family="AT-rich"
              complement(25821..25945)
              /rpt_family="L1MB1"
              25951..26234
              /rpt_family="AluJo"
              complement(26235..26372)
              /rpt_family="L1MB3"
              complement(26746..26798)
              /rpt_family="MLT2B"
              26803..26843
              /rpt_family="(TA)n"
              complement(26846..27250)
              /rpt_family="MLT2B"
              27332..27668
              /rpt_family="THE1C"
              complement(27671..27972)
              /rpt_family="AluSx"
              27976..28014
              /rpt_family="THE1C"
              29127..29427
              /rpt_family="AluJb"
              29444..29716
              /rpt_family="AluJo"
              complement(29995..30356)
              /rpt_family="THE1B"
              30392..30536
              /rpt_family="MIR"
              complement(30630..30933)
              /rpt_family="AluJo"
              30967..31333
              /rpt_family="THE1B"
              31385..31480
              /rpt_family="MIR"
              31575..31870
              /rpt_family="AluSx"
              complement(32080..32358)
              /rpt_family="MLT1B"
              complement(32365..32659)
              /rpt_family="AluJo"
              complement(32672..32841)
              /rpt_family="MLT1B"
              complement(33501..33799)
              /rpt_family="AluSx"
              33866..33991
misc_feature /note="predicted exon, program: gra112exons_human_1.3,
             frame: 1, quality: good, score: 50.000"
              /rpt_family="MIR"
              complement(34438..34738)
repeat_region
```

```

repeat_region /rpt_family="AluSx"
complement(34739..34767)
repeat_region /rpt_family="A7_rich"
complement(34779..34886)
repeat_region /rpt_family="FRAM"
34976..35221
repeat_region /rpt_family="LINE2"
35223..35528
repeat_region /rpt_family="AluSx"
35529..35769
repeat_region /rpt_family="LINE2"
complement(36178..36478)
repeat_region /rpt_family="AluSx"

```

```

Query Match      64.8%  Score 21.4;  DB 9;  Length 81357;
Best Local Similarity 80.6%  Pred. No. 77;
Matches 25;  Conservative 0;  Mismatches 6;  Indels 0;  Gaps 0;

```

Qy 2 tcatctactgcacagcaggagactgtaagt 32

Db 77822 TCATCTACTGAAACATGAGGAAAGG 77852

```

RESULT 13
AC010593 93553 bp DNA linear HTG 18-JUL-2000
LOCUS Homo sapiens chromosome 19 clone CTC-47703, WORKING DRAFT SEQUENCE,
DEFINITION 8 ordered pieces.
ACCESSION AC010593
VERSION AC010593.4 GI:7711470
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT;
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 93553)
REFERENCE DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
AUTHORS 2 (bases 1 to 93553)
DOE Joint Genome Institute.
DIRECT SUBMISSION
Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 6, 2000 this sequence version replaced gi:7690058.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 368687, BC333711
Center clone name: CIT-HSPC_47703
-----
Summary Statistics
Consensus quality: 89161 bases at least Q40
Consensus quality: 92004 bases at least Q30
Consensus quality: 92432 bases at least Q20
Estimated insert size: 99750; agarose-ef estimation
Estimated insert size: 93253; sum-of-contigs estimation
Quality coverage: 9.64 in Q20 bases; agarose-ef estimation
Quality coverage: 10.32 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 2113: contig of 2113 bp in length
* 2114 2213: gap of unknown length

```

```

* 2214 44366: contig of 42153 bp in length
* 44367 44466: gap of unknown length
* 44467 63117: contig of 18651 bp in length
* 63118 63217: gap of unknown length
* 63218 64264: contig of 1047 bp in length
* 64265 64364: gap of unknown length
* 64365 64543: contig of 178 bp in length
* 64543 64642: gap of unknown length
* 64643 65631: contig of 989 bp in length
* 65632 65731: gap of unknown length
* 65732 86549: contig of 20818 bp in length
* 86550 93553: contig of 6904 bp in length.
* Location/Qualifiers
  source          1..93553
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="19"
                  /clone="CTC-47703"
                  /clone_lib="Caltech human BAC library C"
                  /clone_id="19529 g 27970 t 712 others

```

```

BASE COUNT      25617 a 19725 c 19529 g 27970 t 712 others
ORIGIN

```

```

Query Match      64.8%  Score 21.4;  DB 2;  Length 93553;
Best Local Similarity 80.6%  Pred. No. 79;
Matches 25;  Conservative 0;  Mismatches 6;  Indels 0;  Gaps 0;

```

Qy 2 tcatctactgcacagcaggagactgtaagt 32

Db 51464 TCATCTACTGAAACATGAGGAAAGG 51434

```

RESULT 14
AC016587 155221 bp DNA linear PRI 21-JUL-2001
LOCUS Homo sapiens chromosome 19 clone CTD-2652B17, complete sequence.
DEFINITION AC016587
ACCESSION AC016587.8 GI:14993681
VERSION AC016587.8 GI:14993681
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155221)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS 2 (bases 1 to 155221)
DOE Joint Genome Institute.
DIRECT SUBMISSION
Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 155221)
DOE Joint Genome Institute and Stanford Human Genome Center.
DIRECT SUBMISSION
Submitted (25-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 155221)
DOE Joint Genome Institute and Stanford Human Genome Center.
DIRECT SUBMISSION
Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jul 21, 2001 this sequence version replaced gi:14196366.
Draft sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.3.
STS Content:
SHGC-5835 G14094.
Location/Qualifiers

```

source 1. .155221
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CytD-2632B17"

BASE COUNT 41029 a 34040 c 33604 g 46548 t
ORIGIN

Query Match 64.8%; Score 21.4; DB 9; Length 155221;
Best Local Similarity 80.6%; Pred. No. 83;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tctactctgcacagagagactgttaagtgc 32
|||||
Db 149344 TCATCTACTGAAACATGAGAGAGGAAAGGG 149314

RESULT 15

LOCUS 163568 163568 234 bp DNA 1 linear PAT 07-OCT-1997
DEFINITION Sequence 143 from patent US 5663143.
ACCESSION 163568
VERSION 163568.1 GI:2481141
KEYWORDS

SOURCE
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 234)
AUTHORS Ley,A.Charles., Ladner,R.Charles., Guterma,S.Kosow.,
Roberts,B.Lindsay., Markland,W. and Kent,R.Baripault.

TITLE Engineered human-derived kunitz domains that inhibit human
neutrophil elastase

JOURNAL Patent: US 5663143-A 143 02-SEP-1997;

FEATURES
source Location/Qualifiers
1. .234

BASE COUNT 50 a 55 c 62 g 66 t 1 others
ORIGIN /organism="unknown"

Query Match 63.6%; Score 21; DB 6; Length 234;
Best Local Similarity 82.8%; Pred. No. 64;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 tctactgcacagagagactgttaagtgc 33
|||||
Db 116 TCTACTACACTCGAGTATGCTAAGTGC 144

Search completed: May 12, 2002, 20:13:05
Job time: 7167 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 19:41:31 ; Search time 2487.21 Seconds
(without alignments)
341.872 Million cell updates/sec

Title: US-09-554-024-2

Sequence: 1 ggtcccaagaagccagtcgc.....ctgtaagtcgacagagatg 63

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estha:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pla:*
16: em_gss_vlc:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28.4	45.1	629	10	BF384319 602046667
2	28.4	45.1	825	10	BF234717 602028543
3	28.4	45.1	890	10	BF236794 602028092
4	27.4	43.5	652	9	A1573712 u164b04.y
5	26.8	42.5	426	10	W29802 mc24b05.r1
6	26.8	42.5	510	9	A1116411 u102906.y
7	26.8	42.5	560	9	A1037277 u660q11.y
8	26.8	42.5	585	10	BF138224 601781351
9	26.8	42.5	588	10	BG807277 2053-42 M
10	26.8	42.5	588	9	AM255323 ML332 pep
11	26.8	42.5	600	10	BG805295 0614-91 M
12	26.8	42.5	600	10	BG807215 2052-61 M
13	26.8	42.5	600	10	B1989219 4032-71 M
14	26.8	42.5	600	10	B1990353 4064-19 M
15	26.8	42.5	600	10	B1991233 4092-60 M
16	26.8	42.5	608	10	BG906827 T3r11548
17	26.8	42.5	627	9	AA871245 vq32h06.r

C 18	26.8	42.5	736	10	BI332864
C 19	26.8	42.5	754	10	BE915023
C 20	26.8	42.5	759	10	BI554490
C 21	26.8	42.5	774	9	AU080567
C 22	26.8	42.5	778	9	AU080499
C 23	26.8	42.5	822	10	BI833909
C 24	26.8	42.5	836	10	BI220094
C 25	26.8	42.5	910	9	A1266998
C 26	26.8	42.5	1100	10	BM461228
C 27	26.6	42.2	386	9	AA591138
C 28	26.6	42.2	552	12	AO516103
C 29	26.2	41.6	396	10	BG9696508
C 30	26	41.3	565	9	AA098372
C 31	26	41.3	986	10	BF781798
C 32	26	41.3	1708	10	BG685380
C 33	25.8	41.0	485	10	BG892285
C 34	25.8	41.0	595	9	A1195428
C 35	25.8	41.0	651	9	AA475350
C 36	25.2	40.0	496	10	BE496965
C 37	25.2	40.0	584	10	BF473423
C 38	25.2	40.0	600	10	BG803196
C 39	25.2	40.0	602	9	AA771450
C 40	25.2	40.0	661	10	BF570601
C 41	25.2	40.0	676	12	AG105074
C 42	25.2	40.0	715	10	BF578176
C 43	25.2	40.0	717	10	BE586077
C 44	25.2	40.0	760	9	A1182105
C 45	25.2	40.0	778	10	BI330345

ALIGNMENTS

RESULT 1
LOCUS BF384319/c
DEFINITION 602046667/1 NCL_CGAP_L19 Mus musculus cDNA clone IMAGE:419585 5',
ACCESSION BF384319
VERSION BF384319.1 GI:11365624
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 629)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LHAM9532 row: f column: 18
High quality sequence stop: 625.
location/Qualifiers
1. 629
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:419585"
/clone_lib="NCL_CGAP_L19"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCL_CGAP library."

BASE COUNT

169 a 144 c 174 g 142 t


```

/dev_stage="adult"
/1ab_host="DH10B"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
(CACTGTGTG); Site_2: DraIII (CACCATTGTG); 1st strand cDNA
(CACTGTGTG); Site_2: DraIII (CACCATTGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[AGTCGCCCTTTTTTTTTTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [GTGGGCGCTACTGC], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTG, 3' site CACCATTGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCGTCTCTTAACACTCGG and 3' end
primer CGACCTGCAGCTCGACACA."

```

Query Match	42.5%	Score 26.8	DB 9	Length 510
Best Local Similarity	68.5%	Pred No. 1.2e+02		
Matches 37; Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0;

```
Oy 10 aagccagttgccaatcatctactgcaacaggaggaactgtgtaagttgccaggagatg 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 328 AGGCAGAGGCGCTATCTTCAAAATGAACATGAGCATTTGAAAATTCACAGGGGATG 275
```

RESULT	7
AI037277/c	
LOCUS	
DEFINITION	AI037277 560 bp mRNA linear EST 26-JUN-1998
	ud609j1.y1 Sugano mouse liver mlla Mus musculus cDNA clone
IMAGE:	1450340 5' similar to gb:M74570 mouse aldehyde dehydrogenase
II mRNA,	complete cds (MOUSE);, mRNA sequence.

ACCESSION	AI037277	GI:3260968
VERSION	AI037277.1	
KEYWORDS	EST.	

SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE	Manning, A.C., Edgerton, D., Anderson, C.	1967
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.	1980

Geisels, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

TITLE	Waterson, K.
JOURNAL	The Washu-HMI Mouse EST Project Unpublished (1996)

COMMENT *

Contact: Maria M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

This clone is available royalty-free through NLM; contact
IMAGE Consortium (info@image.lnl.gov) for further info
MGI:923656

```
Seq primer: custom primer used
High quality sequence stop: 417.
Location/Qualifiers
FEATURES
source 1. .560
```

```

/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone IMAGE:1450340"
/clone_lib="Sugano mouse liver mla"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: PME185-FL3; Site_1: DraIII
(CACGTGTTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA

```

BASE COUNT	157	a	122	c	145	g	136	t
ORIGIN	<p>was primed with an oligo(dT) primer [ATGGGCGCTTTTTTTTTTTTT]: double-stranded cDNA was ligated to a DraIII adaptor [TCCTGCGCTACGCG], digested and cloned into distinct DraIII sites of the pME18S-PL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCGCTCTAAAGCTGCG and 3' end primer CGACCTGCACTCGACACA."</p>							
Query Match	42.5%	Score 26.8;	DB 9;	Length 560;				
Best Local Similarity	68.5%;	Pred. NO. 1.3e+02;						
Matches 37;	Conservative 0;	Mismatches 17;	Indels 0;	Gaps 0;				
QY	10	aagccagtcgcacatcatctactgcacacaggaagctgtaagtcgcagagatg	63					
		1 1						
Db	510	aaggcagggcccttcttcccaaatgacactgagcattggaaaaattccacaggcgatg	457					

RESULT 8
BFI38224/c
PFI38224
E05 bo
mbna j4noar
ECM

ACCESSION	mRNA sequence.
VERSION	BF138224
KEYWORDS	BF138224.1 GI:10977264
SOURCE	EST.
ORGANISM	house mouse. Mus musculus

REFERENCE
1 (bases 1 to 585)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
KEYWORDS NIH-MGC, MGC, National Institutes of Health, Mammalian Gene Collection (MGC)

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
Procurement: Gilbert Smith, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.

Additional information can be obtained at <http://www.mrg-ncat.gov>

CLONE DISTRIBUTION: MCGAF CLONE DISTRIBUTION
found through the I.M.A.G.E. Consortium/LNLT at:
<http://image.lnlt.gov>

```

Plate: L1AM9245 row: e column: 14
High quality sequence stop: 585.
location/Qualifiers
    source          1..585

```

```

/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:4009333"
/clone_lib="NCI-CCAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; site_1: NotI;
site_2: SalI; transgenic model MNT-1, expression driven by
MNTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dR. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"

```

Query Match	42.5%	Score 26.8	DB 10	Length 585
Best Local Similarity	68.5%	Pred: 1.3e+02		
Matches 37	Conservative	0	Mismatches 17	Indels 0
				Gaps 0

[illegible][illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 18:12:43 ; Search time 2487.21 Seconds
(without alignments)
179.076 Million cell updates/sec

Title: US-09-554-024-1
Perfect score: 1 atcattctgcacacagagactgtaagtgc 33
Sequence: 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
C 1	23	69.7	647 9	AV708999 AV708999
C 2	21.4	64.8	215 10	BG062811 L0958812
C 3	21.4	64.8	495 12	AQ820935 HS_4679_B
C 4	21.4	64.8	1037 11	AK005310 Mus muscu
C 5	21	63.6	479 9	AU084050
C 6	20.8	63.0	512 10	B1045689 MR3-FN020
C 7	20.8	63.0	1708 10	BG685380 602637225
C 8	20.4	61.8	344 10	B1142930
C 9	20.4	61.8	346 10	BF721095
C 10	20.4	61.8	382 9	A1153393
C 11	20.4	61.8	437 10	H29556
C 12	20.4	61.8	470 9	AA071706 mm71a09.r
C 13	20.4	61.8	482 10	W78283 me78c12.r1
C 14	20.4	61.8	496 9	AA16356 ms71n03.r
C 15	20.4	61.8	529 10	BE631498
C 16	20.4	61.8	531 12	BH209770 Sml-47P15
C 17	20.4	61.8	560 9	A1047386 udf5f09.y

C 18	20.4	61.8	504 12	A2507444	A2507444	1M0349N05
C 19	20.4	61.8	753 10	B1149998	B1149998	602849296
C 20	20.4	61.8	755 10	BF783942	BF783942	602109978
C 21	20.4	61.8	786 10	BG780432	BG780432	602791342
C 22	20.4	61.8	888 10	BF784120	BF784120	602110077
C 23	20.4	61.8	2229 11	AK005184	AK005184	Mus muscu
C 24	20.2	61.2	312 10	BF737108	BF737108	PMO-KT004
C 25	20.2	61.2	372 10	BG279687	BG279687	D5603np.f
C 26	20.2	61.2	396 10	BG96508	BG96508	CMO-HT129
C 27	20.2	61.2	408 12	A0645225	A0645225	RPC193-Ec
C 28	20.2	61.2	486 12	A0642870	A0642870	RPC193-Ec
C 29	20.2	61.2	589 9	AM255323	AM255323	ML332 pep
C 30	20.2	61.2	636 12	A0948634	A0948634	Sheared D
C 31	20.2	61.2	639 9	BB650920	BB650920	BB650920
C 32	20.2	61.2	665 10	BF480370	BF480370	LO-2301T3
C 33	20.2	61.2	675 10	BM300861	BM300861	MCA056H02
C 34	20.2	61.2	694 10	BF479025	BF479025	L48-2659T
C 35	20.2	61.2	820 9	BE131115	BE131115	L48-1066T
C 36	20.2	61.2	472 10	BJ036406	BJ036406	BJ036406
C 37	20.2	60.6	521 9	AM128551	AM128551	fe17c12.y
C 38	20.2	60.6	543 9	A1588657	A1588657	fb97n01.y
C 39	20.2	60.6	574 9	AM078159	AM078159	fe25c12.y
C 40	20.2	60.6	716 10	B1326776	B1326776	STR00039
C 41	20.2	60.6	860 10	BE900105	BE900105	601673007
C 42	19.8	60.0	439 12	A2045606	A2045606	T234286b
C 43	19.8	60.0	484 12	A2730423	A2730423	RPC1-24-1
C 44	19.8	60.0	507 12	BH528457	BH528457	BOGMT50TR
C 45	19.8	60.0	534 9	AA719242	AA719242	ah45n07.s

ALIGNMENTS

RESULT 1
LOCUS AV708999 647 bp mRNA linear EST 09-Oct-2000
DEFINITION AV708999 ADC Homo sapiens cDNA clone ADCAXF12 5', mRNA sequence.
ACCESSION AV708999
VERSION AV708999.1 GI:10726264
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukariyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 647)
AUTHORS Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Ou,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
ORGANISM Homo sapiens cDNA ADC clones
COMMENT Unpublished (2000)
CONTACT: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

TITLES

JOURNAL
COMMENT
FEATURES
source
1. 647
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ADCAXF12"
/clone_lib="ADC"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 136 a 172 c 176 g 162 t 1 others
ORIGIN

Query Match 69.7%; Score 23; DB 9; Length 647;
Best Local Similarity 83.9%; Pred. No. 95;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 tcactctgcaacagagagctgtaagt 32
|||||
Db 517 TCCTCTCTGCACACGAGCAGTGTGGAAAGTG 487

RESULT 2
LOCUS BG062811/c 215 bp mRNA linear EST 25-JAN-2001
DEFINITION Musculus cDNA clone L0958B12 5', mRNA sequence.
ACCESSION BG062811
VERSION BG062811.1 GI:12533662
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 215)
Plao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
Construction of long-transcript enriched cDNA libraries from
subclonable amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)

REFERENCE 1
AUTHORS Plao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE Construction of long-transcript enriched cDNA libraries from
subclonable amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)

JOURNAL 21429098
MEDLINE
COMMENT

FEATURES
source
1. 215
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="nlacst:L0958B12-5"
/db_xref="taxon:10090"
/clone="L0958B12"
/clone_1ib="NIA Mouse Newborn Kidney cDNA Library2 (short
)"
/tissue_type="Newborn Kidney"
/dev_stage="Newborn"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is
a short-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]) In brief,
double-stranded cDNAs were synthesized with an Oligo(dT)
primer (Invitrogen); 5'-
pACGAGTCTGATCGCGAGCGCGCCCTTTTCTTTT-3' from 26
ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker LR-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Tag polymerase (Takara) with a primer SalI-4. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 1.5 kb. The library was constructed
by Yulan Plao(NIA)."

BASE COUNT 73 a 46 c 44 g 52 t
ORIGIN

Query Match 64.8%; Score 21.4; DB 10; Length 215;
Best Local Similarity 80.6%; Pred. No. 2.8e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcactctgcaacagagagctgtaagt 32
|||||
Db 202 TCATCTACTGCATCAGCAGACAGTGAAGTG 172

RESULT 3
LOCUS A0820935 495 bp DNA linear GSS 26-AUG-1999
DEFINITION HS_4679_B1_C07.SP6E CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=4679 Col=13 Row=F, DNA sequence.
ACCESSION A0820935
VERSION A0820935.1 GI:5783328
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 495)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shakeri,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
CONTACT: Mahairas GG, Wallace JC, Hood L
UNIVERSITY OF WASHINGTON
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
BAC: 4679 row: F column: 13
Seq primer: T7
Class: BAC ends
High quality sequence stop: 495.
Location/Qualifiers
1. 495
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=4679 Col=13 Row=F"
/clone_1ib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 144 a 95 c 132 g 123 t 1 others
ORIGIN

Query Match 64.8%; Score 21.4; DB 12; Length 495;
Best Local Similarity 80.6%; Pred. No. 3.6e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atcactctgcaacagagagctgtaagt 31
|||||
Db 241 ACCATCAACGTGAACAGAGAGTGTAAATT 271

RESULT 4
LOCUS AK005310 1037 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male cerebellum cDNA, RIKEN full-length enriched
library, clone:11500031D15:unclassified transcript, full insert

Query Match 63.6%: Score 21: DB 9: Length 479;
 Best Local Similarity 82.8%: Pred. No. 5.1e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 atcactctacgcacagagagctgtaagtg 29
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 92 ATCAGCTTACTTCACTGAGCCCTGATATA 120

RESULT 6
 BI045689 512 bp mRNA linear EST 14-JUN-2001
 LOCUS BR3-FN0206-300101-004-c09 FN0206 Homo sapiens CDNA, mRNA sequence.
 DEFINITION BI045689
 ACCESSION BI045689
 VERSION BI045689.1 GI:14452311
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 1 (bases 1 to 512)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brites,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tbl=MR3&ct2=MR3-FN0206-
 300101-004-c09&ct3=2001-01-30&ct4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 25
 High quality sequence stop: 510.
 Location/Qualifiers

FEATURES
 source
 1. 512
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="FN0206"
 /dev_stage="Adult"
 /note="Organ: prostate_normal; Vector: puc18; Site_1: SmaI
 ; Site_2: SmaI; A mini-library was made by cloning
 products derived from ORESTES PCR (U.S. Letters Patent
 Application No. 196,716 - Ludwig Institute for Cancer
 Research) profiles into the puc 18 vector. Reverse
 transcription of tissue mRNA and cDNA amplification were
 performed under low stringency conditions."

BASE COUNT 166 a 144 c 114 g 88 t
 ORIGIN

Query Match 63.0%: Score 20.8: DB 10: Length 512;
 Best Local Similarity 78.1%: Pred. No. 6.2e+02;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 atcactctacgcacagagagctgtaagtg 32
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 461 ATCATCTAATGCACAGCAACAGACAGAGTG 492

RESULT 7

BG685380/c
 LOCUS BG685380 1708 bp mRNA linear EST 01-MAY-2001
 DEFINITION 602637225F1 NIH_MGC_48 Homo sapiens CDNA clone IMAGE:4764942 5',
 mRNA sequence.
 ACCESSION BG685380
 VERSION BG685380.1 GI:13916777
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 1 (bases 1 to 1708)
 NIH-MGC http://imgc.ncbi.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: c9gabs-r@mail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LCM1623 row: a column: 07
 High quality sequence stop: 68.
 Location/Qualifiers

FEATURES
 source
 1. 1708
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:4764942"
 /clone_lib="NIH_MGC_48"
 /tissue_type="primary B-cells from tonsils (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: B-cells; Vector: pORF7; Site_1: XhoI;
 Site_2: EcoRI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by Ling
 Hong in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."

BASE COUNT 486 a 456 c 394 g 372 t
 ORIGIN

Query Match 63.0%: Score 20.8: DB 10: Length 1708;
 Best Local Similarity 78.1%: Pred. No. 8.8e+02;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 atcactctacgcacagagagctgtaagtg 32
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 885 ATCATCTAATGCATAGTAGTAGTGAGTG 854

RESULT 8
 BI142930 344 bp mRNA linear EST 10-JUL-2001
 LOCUS BK76605.Y1 Meloidogyne javanica egg PAMP1 v6 Chiapelli McCarter
 DEFINITION Meloidogyne javanica CDNA 5', mRNA sequence.
 ACCESSION BI142930
 VERSION BI142930.1 GI:14624640
 KEYWORDS EST.
 SOURCE root-knot nematode.
 ORGANISM Meloidogyne javanica

REFERENCE
 AUTHORS
 1 (bases 1 to 344)
 McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
 Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
 Gibbons,M., Rittler,E., Bennett,J., Franklin,C., Tsagarisvill,R.,
 Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe

BASE COUNT	ORIGIN	109 a	74 c	84 g	115 t
Query Match	Best Local Similarity	61.88;	Score 20.4;	DB 9;	Length 382;
Matches	24; Conservative	0;	Mismatches	6;	Indels 0;
Gaps	0;				
Db	173 CATCTACTCCAGCAGCAGCAGCAGCTGATG	144			
RESULT 11	H29556/c				
LOCUS	H29556	437 bp	mRNA	linear	EST 17-JUL-1995
DEFINITION	YMS8d03.s1 Soares infant brain INIB Homo sapiens cDNA clone				
ACCESSION	IMAGE:52880.3				
VERSION	H29556				
KEYWORDS	H29556.1	GI:900466			
SOURCE	EST.				
ORGANISM	human.				
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 437)				
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman				
	M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,				
	Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston				
	,R., Williamson,A., Woldmann,P. and Wilson,R.				
TITLE	The Washu-Merck EST Project				
JOURNAL	Unpublished (1995)				
COMMENT	Contact: Wilson RK				
	Washington University School of Medicine				
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108				
	Tel: 314 286 1800				
	Fax: 314 286 1810				
	Email: estelwatson.wustl.edu				
	Insert Size: 1441				
	High quality sequence stops: 355 Source: IMAGE Consortium, LNL				
	This clone is available royalty-free through LNL; contact the				
	IMAGE Consortium (info@image.llnl.gov) for further information.				
	Insert Length: 1441 Std Error: 0.00				
	Seq primer: Promega -2im13				
	High quality sequence stop: 355.				
FEATURES	Location/Qualifiers				
SOURCE	1. 437				
	/organism="Homo sapiens"				
	/db_xref="GDB:425816"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:52880"				
	/clone_lib="Soares Infant brain INIB"				
	/sex="Female"				
	/dev_stage="73 days post natal"				
	/lab_host="DH10B (ampicillin resistant)"				
	/note="Organ: whole brain; Vector: Lambda BA; Site:1: Not				
	I; Site:2: Hind III; 1st strand cDNA was primed with a Not				
	I - oligo(dT) primer [5				
	ACTGTGAAGATTCGGCCGCGCAGGAATTTTTTTTTTTTTTTT 3'];				

Query Match	Best Local Similarity	Score	DB 10:	Length	437:
Matches	24:	Conservative	0:	Mismatches	6:
				Indels	0:
				Gaps	0:
QY 2	tcactactgcacagagagactgtaagt 31				
Db 379	TCCTATCCAGAAACAGACGAACTGCTAGT 350				
RESULT 12	AA071706	470 bp	mRNA	linear	EST 07-FEB-1997
LOCUS	mm71a09.r1	Stratagene mouse macrophage (#937306)	Mus musculus	CDNA	
DEFINITION	clone IMAGE:533848 5', mRNA sequence.				
ACCESSION	AA071706				
VERSION	AA071706.1	GI:1590052			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 470) Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lag, M., Le, M., Martin, J., Morris, M., Scheinlenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.				
TITLE	The WashU-HMIT Mouse EST Project				
JOURNAL	Unpublished (1996)				
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMIT Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL: contact the IMAGE Consortium (infoimage.llnl.gov) for further information. MGI:320784 Seq primer: -28ml3 rev1 ET from Amersham High quality sequence stop: 444. Location/Qualifiers 1..470 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:533848" /clone_lib="Stratagene mouse macrophage (#937306)" /issue_type="macrophage" /dev_stage="MEH1-3 cell line" /lab_host="SOLR (kanamycin resistant)" /note="Organ: blood; Vector: pBluescript SK-; Site: 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. MEH1-3 cell line. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAAATCGCAGCGAG 3' -3' adaptor sequence: 5' CTCGAGTTTATTTTATTTT 3'"				
BASE COUNT	135 a	86 c	105 g	144 t	
ORIGIN					
Query Match	Best Local Similarity	Score	DB 9:	Length	470:
Matches	24:	Conservative	0:	Mismatches	6:
				Indels	0:
				Gaps	0:
QY 3	catctactgcacacagagagactgtaagt 32				

Db 149 CATCTACTCCAGCAGCAGCAGTGAGTG 120
 RESULT 13
 W78283/c 482 bp mRNA linear EST 21-JUN-1996
 DEFINITION me78c12.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
 W78283
 clone IMAGE:401686 5', mRNA sequence.
 W78283.1 GI:1388746
 EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 482)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL : contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:245454
 Seq primer: EPPRimer
 High quality sequence stop: 267.
 Location/Qualifiers
 1..482
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:401686"
 /clone_lib="Soares mouse embryo NBME13.5 14.5"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTTCACCAATCTGACGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
 State Univ., from 2 1]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT733 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M.Fatima Bonaído."
 BASE COUNT 152 a 98 c 113 g 119 t
 ORIGIN
 Query Match 61.8%; Score 20.4; DB 10; Length 482;
 Best Local Similarity 80.0%; Pred. No. 8.6e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Oy 3 catctactgcaacagagagctgtaagt 32
 Db 292 CATCTACTCCAGCAGCAGCAGTGAGTG 263
 RESULT 14

AA166356/c
 LOCUS AA166356 496 bp mRNA linear EST 16-FEB-1997
 DEFINITION me71n03.r1 Soares mouse 3NBMS Mus musculus cDNA clone IMAGE:617045
 5', mRNA sequence.
 ACCESSION AA166356
 VERSION AA166356.1 GI:1744447
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 496)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL : contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:377869
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 333.
 Location/Qualifiers
 1..496
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:617045"
 /clone_lib="Soares mouse 3NBMS"
 /sex="male"
 /tissue_type="Spleen"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTTCACCAATCTGACGTGGAGCGCGCGCGGTGTGTGTGTGTGTGTGTGT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT733 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through
 three rounds of normalization, and was constructed by
 Bento Soares and M.Fatima Bonaído."
 BASE COUNT 146 a 95 c 108 g 147 t
 ORIGIN
 Query Match 61.8%; Score 20.4; DB 9; Length 496;
 Best Local Similarity 80.0%; Pred. No. 8.7e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Oy 3 catctactgcaacagagagctgtaagt 32
 Db 201 CATCTACTCCAGCAGCAGCAGTGAGTG 172
 RESULT 15
 BE631498/c 529 bp mRNA linear EST 25-AUG-2000
 LOCUS BE631498
 DEFINITION uu60c08.v1 Soares-thymus-2NBMT Mus musculus cDNA clone
 IMAGE:3376334 5', mRNA sequence.
 ACCESSION BE631498
 VERSION BE631498.1 GI:9914186
 KEYWORDS EST.

SOURCE
ORGANISM

house mouse.
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1085938

Seq primer: -40RP from Gibco
High quality sequence stop: 467.

FEATURES
SOURCE

1..529
Location/Qualifiers

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="3376334"
/clone_lib="Soares_Chymus_2NbMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7r3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAGAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7r3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Felima Bonaldo."

BASE COUNT
ORIGIN

147 a 102 c 122 g 158 t

Query Match 61.8%; Score 20.4; DB 10; Length 529;
Best Local Similarity 80.0%; Pred. No. 8.9e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 catctactgcaacagcaggaactgtaagt 32
||||||| ||||| ||||| |||||
Db 173 CATCTACTCCAGCAGCGACAGAGGTGACTG 144

Search completed: May 12, 2002, 19:41:31
Job time: 5328 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OW nucleic - nucleic search, using sw model

Run on: May 12, 2002, 19:41:44 ; Search time 96.6 Seconds
(without alignments)
133.857 Million cell updates/sec

Title: US-09-554-024-1
33
Perfect score: 1 atcatctactgcacacagagagactgtaagtgc 33
Sequence:

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 398110 seqs, 195917913 residues

Total number of hits satisfying chosen parameters: 796220

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/pct_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/us06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/us07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/us08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/us09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/us10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/us60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19.4	58.8	16463	6	US-10-105-299-10527
C 2	19.2	58.2	5053	6	US-10-109-310-20
C 3	19.2	58.2	33000	6	US-10-109-310-18
C 4	18.6	56.4	588	5	US-09-749-7288-4
C 5	18.6	56.4	2005	5	US-09-636-596C-18
C 6	18.6	56.4	9792	5	US-09-636-596C-14
C 7	18.6	56.4	14155	6	US-10-108-605-102
C 8	18.6	56.4	49642	6	US-10-105-299-10868
C 9	18.4	55.8	1511	5	US-09-612-402A-10
C 10	18.4	55.8	1515	5	US-09-612-402A-22
C 11	18.4	55.8	3324	5	US-09-612-402A-24
C 12	18.4	55.8	3354	5	US-09-612-402A-23
C 13	18.4	55.8	4435	5	US-09-612-402A-1
C 14	18.2	55.2	34547	6	US-10-105-299-7482
C 15	18.2	55.2	224	6	US-10-113-872-1613
C 16	18.2	55.2	275	6	US-10-015-219-710
C 17	18.2	55.2	275	6	US-10-015-219-1447
C 18	18.2	55.2	282	6	US-10-015-219-1091
C 19	18.2	55.2	282	6	US-10-015-219-1441
C 20	18.2	55.2	291	6	US-10-015-219-1186
C 21	18.2	55.2	291	6	US-10-015-219-1450
C 22	18.2	55.2	2650	6	US-10-106-698-2090
C 23	18.2	55.2	3347	6	US-10-000-773A-16
C 24	18.2	55.2	6728	1	PCT-US02-08253-250
C 25	18.2	55.2	6739	6	US-10-116-802-237
C 26	17.8	53.9	96	5	US-09-539-331D-23554

C 27	17.8	53.9	98	5	US-09-540-210B-6858	Sequence 6858, Ap
C 28	17.8	53.9	153	5	US-09-540-210B-29815	Sequence 29815, A
C 29	17.8	53.9	407	6	US-10-106-698-2497	Sequence 2497, Ap
C 30	17.8	53.9	460	5	US-09-539-331D-35624	Sequence 35624, A
C 31	17.8	53.9	1723	6	US-10-105-299-1536	Sequence 1536, Ap
C 32	17.8	53.9	2704	6	US-10-116-512-3	Sequence 3, Appl
C 33	17.8	53.9	3179	1	PCT-US02-09135-134	Sequence 134, Appl
C 34	17.8	53.9	3179	1	PCT-US02-09105-191	Sequence 191, App
C 35	17.8	53.9	3179	1	PCT-US02-09188-390	Sequence 390, App
C 36	17.8	53.9	3179	1	PCT-US02-09257-257	Sequence 257, App
C 37	17.8	53.9	3179	1	PCT-US02-09370-404	Sequence 404, App
C 38	17.8	53.9	3179	1	PCT-US02-09922-253	Sequence 253, App
C 39	17.8	53.9	3179	6	US-10-105-299-3099	Sequence 3099, Ap
C 40	17.8	53.9	3258	5	US-09-540-209B-8	Sequence 8, Appl
C 41	17.8	53.9	4015	6	US-10-106-698-1018	Sequence 1018, Ap
C 42	17.8	53.9	4852	6	US-10-105-299-10103	Sequence 10103, A
C 43	17.6	53.3	151	5	US-09-685-209A-68	Sequence 68, Appl
C 44	17.6	53.3	168	5	US-09-975-254-28871	Sequence 28871, A
C 45	17.6	53.3	223	5	US-09-540-210B-24915	Sequence 24915, A

ALIGNMENTS

RESULT 1
US-10-105-299-10527/c
Sequence 10527, Application US/10105299
GENERAL INFORMATION:
APPLICANT: Rosen, et. al
TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: PS950
CURRENT APPLICATION NUMBER: US/10/105,299
CURRENT FILING DATE: 2002-03-26
NUMBER OF SEQ ID NOS: 15197
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10527
LENGTH: 16463
TYPE: DNA
ORGANISM: Homo sapiens
US-10-105-299-10527

Query Match 58.8%; Score 19.4; DB 6; Length 16463;
Best Local Similarity 79.3%; Pred. No. 20;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 atcatctactgcacacagagagactgta 29
Db 10709 ATCATCTACTGCTGTACAGAGGCTGTGAA 10681

RESULT 2
US-10-109-310-20
Sequence 20, Appl
GENERAL INFORMATION:
APPLICANT: Wisconsin Alumni Research Foundation
APPLICANT: Hutchinson, Charles R.
APPLICANT: Kennedy, Jonathan n.m.1
APPLICANT: Park, Choonsok n.m.1
TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS
FILE REFERENCE: 960296.95718
CURRENT APPLICATION NUMBER: US/10/109,310
CURRENT FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: US/09/215,694
PRIOR FILING DATE: 1999-12-18
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
LENGTH: 5053
TYPE: DNA
ORGANISM: Aspergillus terreus
US-10-109-310-20

Query Match	58.2%	Score 19.2;	DB 6;	Length 5053;
Best Local Similarity	75.0%;	Pred. No. 20;		
Matches 24; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0

```

RESULT 3
US-10-109-310-18/C
; Sequence 18 Application US/10109310
; GENERAL INFORMATION:
; APPLICANT: Wisconsin Alumni Research Foundation
; APPLICANT: Hutchinson, Charles R.
; APPLICANT: Kennedy, Jonathan n.m.1
; APPLICANT: Park, Cheonseok n.m.1
; TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS
; FILE REFERENCE: 960296 95718
; CURRENT APPLICATION NUMBER: US/10/109,310
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US/09/215,694
; PRIOR FILING DATE: 1999-12-18
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 33000
; TYPE: DNA
; ORGANISM: Aspergillus terreus
; US-10-109-310-18

```

RESULT 4
US-09-749-728B-4
Sequence 4, Application US/09749728B
GENERAL INFORMATION:
APPLICANT: Umezawa, Akihito
APPLICANT: Hata, Jun-ichi
APPLICANT: Fukuda, Keiichi
APPLICANT: Ogawa, Satoshi
APPLICANT: Sakurada, Kazuhiro
APPLICANT: Cojo, Satoshi
APPLICANT: Yamada, Yoji
TITLE OF INVENTION: THE CELL HAVING THE POTENTIALITY OF DIFFERENTIATION INTO CARDIOMYOT
FILE REFERENCE: 00766.000043
CURRENT APPLICATION NUMBER: US/09/749,728B
CURRENT FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: H11-372826
PRIOR FILING DATE: 1995-12-28
PRIOR APPLICATION NUMBER: PCT-JP00-01148
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: PCT-JP00-07741
PRIOR FILING DATE: 2000-11-02
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn Ver.2.0
SEQ ID NO 4

US-09-749-728B-4

Query Match	56.48;	Score 18.6;	DB 5;	Length 588;
Best Local Similarity	72.7%;	Pred. NO. 26;		
Matches 24;	Conservative 0;	Mismatches 9;	Indels 0;	Gaps 0;

```

RESULT 5
US-09-636-596C-18
: Sequence 18, Application US/09636596C
: GENERAL INFORMATION:
: APPLICANT: CANFIELD, WILLIAM
: TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCNAGASE OF THE LYSOSOMAL TARGETING PATH
: FILE REFERENCE: 10529-0001-77
: CURRENT APPLICATION NUMBER: US/09/636,596C
: CURRENT FILING DATE: 2000-08-10
: PRIOR APPLICATION NUMBER: 60/153,831
: PRIOR FILING DATE: 1999-09-14
: NUMBER OF SEQ ID NOS: 52
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 18
: LENGTH: 2005
: TYPE: DNA
: ORGANISM: Mus musculus
US-09-636-596C-18

```

```

RESULT      6
US-09-636-596C-14
: Sequence 14, Application US/09636596C
: GENERAL INFORMATION:
: APPLICANT: CANFIELD, WILLIAM
: TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANASE OF THE LYSOSOMAL TARGETING PATH
: FILE REFERENCE: 10929-0001-77
: CURRENT APPLICATION NUMBER: US/09/636,596C
: CURRENT FILING DATE: 2000-08-10
: PRIOR APPLICATION NUMBER: 60/153,831
: PRIOR FILING DATE: 1999-09-14
: NUMBER OF SEQ ID NOS: 52
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 14
: LENGTH: 9792
: TYPE: DNA
: ORGANISM: Mus musculus
: US-09-636-596C-14

```

```

;
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: (1)..(591)
;
; RESULT 7
; US-10-108-605-102
; Sequence 102, Application US/10108605

```



```

: GENERAL INFORMATION:
: APPLICANT: Broadus, Julie
: APPLICANT: Stem, Lynn
: APPLICANT: Bachmann, Jane
: APPLICANT: Kandari, Kim
: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
: TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
: FILE REFERENCE: 3113B
: CURRENT APPLICATION NUMBER: US/10/108,605
: CURRENT FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: US 09/761,142
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/176,418
: PRIOR FILING DATE: 2000-01-14
: NUMBER OF SEQ ID NOS: 361
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 102
: LENGTH: 14155
: TYPE: DNA
: ORGANISM: Drosophila melanogaster
: US-10-108-605-102

```

Query Match	56.4%	Score 18.6;	DB 6;	length 14155;
Best Local Similarity	72.7%;	Pred. NO. 45;		
Matches 24; Conservative	0;	Mismatches 9;	Indels 0;	Gaps 0;

QY 1 atcatctactgtcaacagagagtgtgtaagtgc 33
 | | | | | | | | | | | | | |
Db 3548 accaacgactgtcaacqgtaccacctgagaaqtgc 3580

```

RESULT      8
US-10-105-299-10868
; Sequence 10868, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: P9350
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10868
; LENGTH: 49642
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-105-299-10868

```

Query Match	56.4%	Score 18.6;	DB 6;	Length 49642;
Best Local Similarity	84.08;	Pred. No. 56;		
Matches 21: Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

```

QY      1  atcatctactgcacacagagactg  25
          |||||
Db 23510 atcatctactgtagtagtagactg  23534

```

```

RESULT      9
US-09-612-402A-10
: Sequence 10, Application US/09612402A
: GENERAL INFORMATION:
: APPLICANT: Jackson, W. James
: APPLICANT: Pace, John
: TITLE OF INVENTION: Chlamydia Protein, Gene Sequence and Uses Thereof
: FILE REFERENCE: 7969-086-999
: CURRENT APPLICATION NUMBER: US/09/612,402A
: CURRENT FILING DATE: 2002-03-29
: PRIOR APPLICATION NUMBER: 08/942,596
: PRIOR FILING DATE: 1997-10-02
: NUMBER OF SEQ. ID NOS: 43

```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1511
; TYPE: DNA
; ORGANISM: Chlamydia sp.
US-09-612-402A-10

```

Query Match	55.8%	Score 18.4	DB 5	Length 1511
Best Local Similarly	78.6%	Pred. No. 37		
Matches 22	Conservative	0	Mismatches 6	Indels 0
				Gaps 0

QY	1	atcatctactgcaacagaggagtgtta	28
Dd	696	atcatctacttcaacagaagatccagta	723

```

RESULT 10
US-09-612-402A-22
: Sequence 22, Application US/09612402A
: GENERAL INFORMATION:
: APPLICANT: Jackson, W. James
: TITLE OF INVENTION: Chlamydia Protein, Gene Sequence and Uses Thereof
: FILE REFERENCE: 7969-086-999
: CURRENT APPLICATION NUMBER: US/09/612,402A
: CURRENT FILING DATE: 2002-03-29
: PRIOR APPLICATION NUMBER: 08/942,596
: PRIOR FILING DATE: 1997-10-02
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 22
: LENGTH: 1515
: TYPE: DNA
: ORGANISM: Chlamydia sp.
US-09-612-402A-22

```

Query Match	55.8%	Score	18.4	DB	5	Length	1515
Best Local	Similarly	Pred. No.	37				
Matches	22; Conservative	Mismatches	6			Indels	0
						Gaps	0

```
QY      1 atcatctactgtcaacagagagactgta 28
          |||||
Db      696 atcatctacttcaacagagaatccagta 723
```

```

RESULT 11
US-09-612-402A-24
; Sequence 24, Application US/09612402A
; GENERAL INFORMATION:
; APPLICANT: Jackson, W. James
; APPLICANT: Pace, John
; TITLE OF INVENTION: Chlamydia Protein, Gene Sequence and Uses Thereof
; FILE REFERENCE: 7969-086-999
; CURRENT APPLICATION NUMBER: US/09/612,402A
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 08/942,596
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 3324
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant Expression Vec
US-09-612-402A-24

```

Query Match	55.88;	Score 18.4;	DB 5;	Length 3324;
Best Local Similarity	78.68;	Pred. No. 43;		
Matches 22; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 18:59:44 ; Search time 4348.96 seconds

(without alignments)
164.166 Million cell updates/sec

Title: US-09-554-024-1

Perfect score: 33
1 atcattactgcacagagagactgtaagtgc 33

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Database : Listing first 45 summaries

```
Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/US06005_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06006_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07007_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08008_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US08009_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US08010_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US08011_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US08012_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US08013_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US08014_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US08015_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US08016_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US08017_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US08018_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US08019_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US08020_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US08021_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US08022_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US08023_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US08024_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US08025_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US08026_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US08027_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US08028_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US08029_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US08030_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US08031_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US08032_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US08033_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US08034_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US08035_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US08036_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US08037_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US08038_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US08039_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US08040_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US08041_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US08042_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US08043_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US08044_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US08045_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US08046_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US08047_COMB.seq:*
```

Result				SUMMARIES			
No.	Score	Match	Length	ID	Description		
1	33	100.0	33	US-09-554-024-1	Sequence 1, Appl		
2	33	100.0	63	US-09-554-024-2	Sequence 2, Appl		
3	33	100.0	87	US-09-554-024-9	Sequence 9, Appl		
4	33	100.0	98	US-09-554-024-3	Sequence 3, Appl		
5	33	100.0	197	US-09-554-024-5	Sequence 5, Appl		
6	22.4	67.9	3102	US-08-265-534A-5	Sequence 5, Appl		
7	22.4	67.9	3102	US-08-456-341-5	Sequence 5, Appl		
8	22.4	67.9	3102	US-08-457-001-5	Sequence 5, Appl		
9	22.4	67.9	3102	US-08-457-543-5	Sequence 5, Appl		
10	22.4	67.9	3102	US-08-457-543-5	Sequence 5, Appl		
11	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
12	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
13	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
14	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
15	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
16	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
17	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
18	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
19	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
20	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
21	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
22	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
23	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
24	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
25	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
26	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
27	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
28	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
29	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
30	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
31	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
32	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
33	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
34	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
35	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
36	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
37	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
38	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
39	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
40	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
41	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
42	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		
43	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap		

```
32 20.2 61.2 1244 62 US-60-234-690-977 Sequence 977, App
33 20.2 61.2 1244 71 US-60-323-843-977 Sequence 977, App
34 20.2 61.2 2178 18 US-09-493-276-32 Sequence 32, Appl
35 20.2 61.2 2268 26 US-09-663-779-976 Sequence 576, Appl
36 20.2 61.2 106901 63 US-60-243-733-1 Sequence 1, Appl1
37 20.2 60.6 225 18 US-09-521-640-237893 Sequence 237893,
38 20.2 60.6 419 18 US-09-465-877-1396 Sequence 1396, Ap
39 20.2 60.6 419 34 US-09-906-555-1396 Sequence 1396, Ap
40 20.2 60.6 427 17 US-09-399-720-9839 Sequence 9839, Ap
41 20.2 60.6 427 34 US-09-921-378-9839 Sequence 9839, Ap
42 20.2 60.6 434 18 US-09-465-877-2338 Sequence 2338, Ap
43 20.2 60.6 434 34 US-09-906-555-2338 Sequence 2338, Ap
44 20.2 60.6 444 18 US-09-465-877-3009 Sequence 3009, Ap
45 20.2 60.6 444 19 US-09-519-705-5860 Sequence 5860, Ap
```

ALIGNMENTS

```
RESULT 1
US-09-554-024-1
```

```
; Sequence 1, Application US/09554024
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Freysinet, Georges
```

```
; APPLICANT: Derose, Richard
```

```
; APPLICANT: Hoffman, Jules
```

```
; TITLE OF INVENTION: Gene Coding for Thanatin, Vector
```

```
; TITLE OF INVENTION: Containing Same and Resulting Transformed Disease-Resistant
```

```
; TITLE OF INVENTION: Plants
```

```
; FILE REFERENCE: A33207-PCT-USA
```

```
; CURRENT APPLICATION NUMBER: US/09/554,024
```

```
; PRIOR FILING DATE: 2000-05-08
```

```
; PRIOR APPLICATION NUMBER: PCT/FR98/02375
```

```
; PRIOR FILING DATE: 1998-11-06
```

```
; PRIOR APPLICATION NUMBER: FR 97/14,263
```

```
; PRIOR FILING DATE: 1997-11-07
```

```
; NUMBER OF SEQ ID NOS: 14
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 1
```

```
; LENGTH: 33
```

```
; TYPE: DNA
```

```
; ORGANISM: Psodius maculiventis
```

```
; FEATURE:
```

```
; NAME/KEY: CDS
```

```
; LOCATION: (1)...(33)
```

```
US-09-554-024-1
```

```
Query Match 100.0%; Score 33; DB 22; Length 33;
```

```
Best Local Similarity 100.0%; Pred. No. 0.00075;
```

```
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 atcatctactgcacagaggagactgtaagtgc 33
Db 1 atcatctactgcacagaggagactgtaagtgc 33
```

```
RESULT 2
US-09-554-024-2
```

```
; Sequence 2, Application US/09554024
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Freysinet, Georges
```

```
; APPLICANT: Derose, Richard
```

```
; APPLICANT: Hoffman, Jules
```

```
; TITLE OF INVENTION: Gene Coding for Thanatin, Vector
```

```
; TITLE OF INVENTION: Containing Same and Resulting Transformed Disease-Resistant
```

```
; TITLE OF INVENTION: Plants
```

```
; FILE REFERENCE: A33207-PCT-USA
```

```
; CURRENT APPLICATION NUMBER: US/09/554,024
```

```
; PRIOR FILING DATE: 2000-05-08
```

```
; PRIOR APPLICATION NUMBER: PCT/FR98/02375
```

```
; PRIOR FILING DATE: 1998-11-06
```

```
; PRIOR APPLICATION NUMBER: FR 97/14,263
```

```
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 63
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(63)
; OTHER INFORMATION: Derived from Psodius maculiventis
US-09-554-024-2
```

```
Query Match 100.0%; Score 33; DB 22; Length 63;
```

```
Best Local Similarity 100.0%; Pred. No. 0.00085;
```

```
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 atcatctactgcacagaggagactgtaagtgc 33
Db 22 atcatctactgcacagaggagactgtaagtgc 54
```

```
RESULT 3
```

```
US-09-554-024-9/c
```

```
; Sequence 9, Application US/09554024
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Freysinet, Georges
```

```
; APPLICANT: Derose, Richard
```

```
; APPLICANT: Hoffman, Jules
```

```
; TITLE OF INVENTION: Gene Coding for Thanatin, Vector
```

```
; TITLE OF INVENTION: Containing Same and Resulting Transformed Disease-Resistant
```

```
; TITLE OF INVENTION: Plants
```

```
; FILE REFERENCE: A33207-PCT-USA
```

```
; CURRENT APPLICATION NUMBER: US/09/554,024
```

```
; PRIOR FILING DATE: 2000-05-08
```

```
; PRIOR APPLICATION NUMBER: PCT/FR98/02375
```

```
; PRIOR FILING DATE: 1998-11-06
```

```
; PRIOR APPLICATION NUMBER: FR 97/14,263
```

```
; PRIOR FILING DATE: 1997-11-07
```

```
; NUMBER OF SEQ ID NOS: 14
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 9
```

```
; LENGTH: 87
```

```
; TYPE: DNA
```

```
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
```

```
; OTHER INFORMATION: Synthetic oligonucleotide
```

```
US-09-554-024-9
```

```
Query Match 100.0%; Score 33; DB 22; Length 87;
```

```
Best Local Similarity 100.0%; Pred. No. 0.0009;
```

```
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 atcatctactgcacagaggagactgtaagtgc 33
Db 77 ATCATCTACTGCACAGAGGAGACTGCTAAGTGC 45
```

```
RESULT 4
US-09-554-024-3
```

```
; Sequence 3, Application US/09554024
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Freysinet, Georges
```

```
; APPLICANT: Derose, Richard
```

```
; APPLICANT: Hoffman, Jules
```

```
; TITLE OF INVENTION: Gene Coding for Thanatin, Vector
```

```
; TITLE OF INVENTION: Containing Same and Resulting Transformed Disease-Resistant
```

```
; TITLE OF INVENTION: Plants
```

```
; FILE REFERENCE: A33207-PCT-USA
```

```
; CURRENT APPLICATION NUMBER: US/09/554,024
```

```
; PRIOR FILING DATE: 2000-05-08
```

;; PRIOR APPLICATION NUMBER: PCT/FR98/02375
;; PRIOR FILING DATE: 1998-11-06
;; PRIOR APPLICATION NUMBER: FR 97/14,263
;; PRIOR FILING DATE: 1997-11-07
;; NUMBER OF SEQ ID NOS: 14
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 3
;; LENGTH: 98
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Derived from Psodius maculiventris
;; NAME/KEY: CDS
;; LOCATION: (1)...(63)
US-09-554-024-3

Query Match 100.0%; Score 33; DB 22; Length 98;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcattactgcacacagagagactggtagtgc 33
|||
Db . 22 atcattactgcacacagagagactggtagtgc 54

RESULT 5
US-09-554-024-5
;; Sequence 5, Application US/09554024
;; GENERAL INFORMATION:
;; APPLICANT: Freyssinet, Georges
;; APPLICANT: Derose, Richard
;; APPLICANT: Hoffman, Jules
;; TITLE OF INVENTION: Gene Coding for Thbanatin, Vector
;; TITLE OF INVENTION: Containing Same and Resulting Transformed Disease-Resistant
;; FILE REFERENCE: A33207-PCT-USA
;; CURRENT APPLICATION NUMBER: US/09/554,024
;; CURRENT FILING DATE: 2000-05-08
;; PRIOR APPLICATION NUMBER: PCT/FR98/02375
;; PRIOR FILING DATE: 1998-11-06
;; PRIOR APPLICATION NUMBER: FR 97/14,263
;; PRIOR FILING DATE: 1997-11-07
;; NUMBER OF SEQ ID NOS: 14
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 5
;; LENGTH: 197
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Derived from Psodius maculiventris
;; NAME/KEY: CDS
;; LOCATION: (12)...(164)
US-09-554-024-5

Query Match 100.0%; Score 33; DB 22; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcattactgcacacagagagactggtagtgc 33
|||
Db 123 atcattactgcacacagagagactggtagtgc 155

RESULT 6
US-08-265-534A-5
;; Sequence 5, Application US/08265534A
;; GENERAL INFORMATION:
;; APPLICANT: Yaver, Debbie Sue
;; APPLICANT: Xu, Feng
;; APPLICANT: Dalboege, Henrik
;; APPLICANT: Schneider, Palle

;; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
;; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
;; NUMBER OF SEQUENCES: 10
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Novo Nordisk of North America, Inc.
;; STREET: 405 Lexington Avenue, Suite 6400
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10174-6401
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25 (PPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/265,534A
;; FILING DATE: 24-JUN-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Lowney, Karen A.
;; REGISTRATION NUMBER: 31,274
;; REFERENCE/DOCKET NUMBER: 4185,000-US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212 867 0123
;; TELEFAX: 212 867 0298
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3102 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; ORIGINAL SOURCE:
;; ORGANISM: Polyporus pinistius
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 666..720
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 790..845
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 1125..1182
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 1390..1450
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 1607..1661
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 1863..1918
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 1976..2025
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 2227..2285
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 2403..2458
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 2576..2627
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: join (665..721, 789..846, 1124..1183, 1389..1451,
;; LOCATION: 1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
;; LOCATION: 2575..2628).

US-08-265-534A-5

Query Match 67.9%: Score 22.4; DB 6; Length 3102;
Best Local Similarity 81.2%; Pred. No. 72;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcatctactgcaacagagagactgtaagtgc 33
DB 2623 TCATCTTTGCAACAGAGAGACTGGAACAAGC 2654

RESULT 7
US-08-456-341-5
; Sequence 5, Application US/08456341
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Xu, Feng
; APPLICANT: Dalboge, Henrik
; APPLICANT: Schneider, Palie
; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,341
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/265,534
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4185,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0298
; TELEFAX: 212 867 0123
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Polyporus plustus
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 666..720
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 790..845
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1125..1182
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1390..1450
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1607..1661
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1863..1918

; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1976..2025
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 2227..2285
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 2403..2458
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 2576..2627
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (665..721, 789..846, 1124..1183, 1389..1451,
; LOCATION: 1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
; LOCATION: 2575..2628).
US-08-456-341-5

Query Match 67.9%: Score 22.4; DB 8; Length 3102;
Best Local Similarity 81.2%; Pred. No. 72;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcatctactgcaacagagagactgtaagtgc 33
DB 2623 TCATCTTTGCAACAGAGAGACTGGAACAAGC 2654

RESULT 8
US-08-457-001-5
; Sequence 5, Application US/08457001
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Xu, Feng
; APPLICANT: Dalboge, Henrik
; APPLICANT: Schneider, Palie
; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,001
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/265,534
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4185,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0298
; TELEFAX: 212 867 0123
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```

; ORIGINAL SOURCE:
; ORGANISM: Polyporus pinsitus
; FEATURE:
; NAME/KEY: intron
; LOCATION: 666..720
; FEATURE:
; NAME/KEY: intron
; LOCATION: 790..845
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1125..1182
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1390..1450
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1607..1661
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1863..1918
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1976..2025
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2227..2285
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2403..2458
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2576..2627
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (665..721, 789..846, 1124..1183, 1389..1451,
; LOCATION: 1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
; LOCATION: 2575..2628).
;
US-08-457-001-5
;
Query Match 67.9% Score 22.4; DB 8; Length 3102;
Best Local Similarity 81.2%; Pred. No. 72;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 tcactctacgcaacagagactgtaagtgc 33
||||| ||||||||| |||
Db 2623 TCATCTTTTCACACGAGGACTGGAACAAGC 2654

RESULT 9
US-08-457-349-5
; Sequence 5, Application US/08457349
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Xu, Feng, Debbie Sue
; APPLICANT: Dalboege, Henrik
; APPLICANT: Schneider, Palie
; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/08/457,349
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/265,534
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowmey, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4185,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Polyporus pinsitus
; FEATURE:
; NAME/KEY: intron
; LOCATION: 666..720
; FEATURE:
; NAME/KEY: intron
; LOCATION: 790..845
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1125..1182
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1390..1450
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1607..1661
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1863..1918
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1976..2025
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2227..2285
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2403..2458
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2576..2627
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (665..721, 789..846, 1124..1183, 1389..1451,
; LOCATION: 1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
; LOCATION: 2575..2628).
;
US-08-457-349-5
;
Query Match 67.9% Score 22.4; DB 8; Length 3102;
Best Local Similarity 81.2%; Pred. No. 72;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 tcactctacgcaacagagactgtaagtgc 33
||||| ||||||||| |||
Db 2623 TCATCTTTTCACACGAGGACTGGAACAAGC 2654

RESULT 10
US-08-457-543-5
; Sequence 5, Application US/08457543
; GENERAL INFORMATION:
```

1	APPLICANT:	Yaver, Debbie Sue
2	APPLICANT:	Xu, Feng
3	APPLICANT:	Dalboge, Henrik
4	APPLICANT:	Schneider, Palle
5	TITLE OF INVENTION:	PURIFIED POLYPORUS LACCASES AND
6	TITLE OF INVENTION:	NUCLEIC ACIDS ENCODING SAME
7	NUMBER OF SEQUENCES:	10
8	CORRESPONDENCE ADDRESS:	
9	ADDRESSEE:	Novo Nordisk of North America, Inc.
10	STREET:	405 Lexington Avenue, Suite 6400
11	CITY:	New York
12	STATE:	New York
13	COUNTRY:	U.S.A.
14	ZIP:	10174-6401
15	COMPUTER READABLE FORM:	
16	MEDIUM TYPE:	Floppy disk
17	COMPUTER:	IBM PC compatible
18	OPERATING SYSTEM:	PC-DOS/MS-DOS
19	SOFTWARE:	Patentin Release #1.0, Version #1.25 (EPO)
20	CURRENT APPLICATION DATA:	
21	APPLICATION NUMBER:	US/08/457,543
22	FILING DATE:	01-JUN-1995
23	CLASSIFICATION:	435
24	PRIOR APPLICATION DATA:	
25	APPLICATION NUMBER:	08/265,534
26	FILING DATE:	24-JUN-1994
27	ATTORNEY/AGENT INFORMATION:	
28	NAME:	Lowney, Karen A.
29	REGISTRATION NUMBER:	31,274
30	REFERENCE/DOCKET NUMBER:	4185.000-US
31	TELECOMMUNICATION INFORMATION:	
32	TELEPHONE:	212 867 0123
33	TELEFAX:	212 867 0298
34	INFORMATION FOR SEQ ID NO:	5:
35	SEQUENCE CHARACTERISTICS:	
36	LENGTH:	3102 base pairs
37	TYPE:	nucleic acid
38	STRANDEDNESS:	double
39	TOPOLOGY:	linear
40	MOLECULE TYPE:	DNA (genomic)
41	ORIGINAL SOURCE:	
42	ORGANISM:	Polyporus pilsitius
43	FEATURE:	
44	NAME/KEY:	Inttron
45	LOCATION:	666..720
46	FEATURE:	
47	NAME/KEY:	Inttron
48	LOCATION:	790..845
49	FEATURE:	
50	NAME/KEY:	Inttron
51	-LOCATION:	1125..1182
52	FEATURE:	
53	NAME/KEY:	Inttron
54	LOCATION:	1390..1450
55	FEATURE:	
56	NAME/KEY:	Inttron
57	LOCATION:	1607..1661
58	FEATURE:	
59	NAME/KEY:	Inttron
60	LOCATION:	1863..1918
61	FEATURE:	
62	NAME/KEY:	Inttron
63	LOCATION:	1976..2025
64	FEATURE:	
65	NAME/KEY:	Inttron
66	LOCATION:	2227..2285
67	FEATURE:	
68	NAME/KEY:	Inttron
69	LOCATION:	2403..2458
70	FEATURE:	
71	NAME/KEY:	Inttron
72	LOCATION:	2576..2627
73	FEATURE:	

```

;      NAME/KEY:   CDS
;      LOCATION:   join (665.,721.,789.,846.,1124.,1183.,1389.,1451,
;      LOCATION:   1606.,1662.,1862.,1919.,1975.,2026.,2266.,2286.,2402.,2459
;      LOCATION:   2575.,2628).
;
US-08-457-543-5

```

Query Match	67.9%;	Score 22.4;	DB 8;	Length 3102;
Best Local Similarity	81.2%;	Pred. No. 72;		
Matches 26; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

QY	2	tcatctactgcacacagggagactgtgtaagtgc	33
Db	2623	TCATCTTTTGCAACAGGAGGACTGGAACAAGC	2654

RESULT 11
US-09-248-796-8333/C
Sequence 8333, Application US/09248796

```

? APPLICANT: Keith Weinstein at al
? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBID
? TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
? FILE REFERENCE: 107196.132
? CURRENT APPLICATION NUMBER: US/09/248.796
? CURRENT FILING DATE: 1999-02-12
? NUMBER OF SEQ ID NOS: 28206
? SEQ ID NO 8333
? LENGTH: 1230
? TYPE: DNA
? ORGANISM: Candida albicans
US-09-248-796-8333

```

	Query Match	66.7%	Score 22;	DB 16;	Length 1230;
	Best Local Similarity	83.3%	Pred. No. 89;		
Matches	22; Conservative	0;	Mismatches	5;	Indels 0; Gaps 0;
OY	3 catctactgcacagcaggactggttaagt	32			
Db	625 CATGTACTGCACACAGGATGCTCGTCATGTG	596			

```

RESULT      12
US-60-096-409-8333/C
; Sequence 8333, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 8333
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Candida albicans
US-60-096-409-8333

```

Query Match	66.7%;	Score 22;	DB 48;	Length 1230;
Best Local Similarity	83.3%;	Pred. NO. 89;		
Matches 25; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

OY		3	catctactgcacagaggagtgtgaatg	32
Db		625	CATGTACTGCACACAGGATGACTCGTCATG	596

RESULT 13
US-09-620-392-35280/c
: Sequence 35280, Application US/09620392


```
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
; FILE REFERENCE: 38-21(51237)E
; CURRENT APPLICATION NUMBER: US/09/620,392
; CURRENT FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 69652
; SEQ ID NO 35280
; LENGTH: 12414
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-620-392-35280

Query Match          64.8%; Score 21.4; DB 24; Length 12414;
Best Local Similarity 80.6%; Pred. No. 2.6e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atcattactgcacagagagactgtaagt 31
    ||| ||||| ||||| ||||| |||||
DB 3272 ATATATACTGCAACAGAGGCTGTTAAGT 3242

RESULT 14
US-09-620-392-67364/C
; Sequence 67364, Application US/09620392
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
; FILE REFERENCE: 38-21(51237)E
; CURRENT APPLICATION NUMBER: US/09/620,392
; CURRENT FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 69652
; SEQ ID NO 67364
; LENGTH: 21712
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-620-392-67364

Query Match          64.8%; Score 21.4; DB 24; Length 21712;
Best Local Similarity 80.6%; Pred. No. 2.9e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atcattactgcacagagagactgtaagt 31
    ||| ||||| ||||| ||||| |||||
DB 3359 ATATATACTGCAACAGAGGCTGTTAAGT 3329

RESULT 15
US-09-702-134-7465/C
; Sequence 7465, Application US/09702134
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Plant Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-21(51237)F
; CURRENT APPLICATION NUMBER: US/09/702,134
; CURRENT FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 52202
; SEQ ID NO 7465
; LENGTH: 66687
; TYPE: DNA
```

```
; ORGANISM: Oryza sativa
US-09-702-134-7465

Query Match          64.8%; Score 21.4; DB 28; Length 66687;
Best Local Similarity 80.6%; Pred. No. 3.6e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atcattactgcacagagagactgtaagt 31
    ||| ||||| ||||| ||||| |||||
DB 3349 ATATATACTGCAACAGAGGCTGTTAAGT 3319
```

Search completed: May 12, 2002, 21:33:35
Job time: 9231 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 18:14:43 ; Search time 76.1 Seconds
(without alignments)
106.516 Million cell updates/sec

Title: US-09-554-024-1
Sequence: 1 atcattctgcaacagagagactgttaagtc 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2-6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2-6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2-6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2-6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2-6/ptodata/2/ina/PTOS.COMB.seq:*
6: /cgn2-6/ptodata/2/ina/Dackfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length DB	ID	Description
1	22.4	67.9	3102	1	US-08-462-484-5
2	22.4	67.9	3102	1	US-08-441-147-5
3	22.4	67.9	3102	5	PCT-US95-07536-5
4	21	63.6	234	1	US-08-358-160-143
5	20.2	61.2	18073	4	US-09-078-294-12
6	19.8	60.0	180	1	US-08-358-160-145
7	18.8	57.0	824	1	US-08-158-353-1
8	18.6	56.4	330	5	PCT-US91-02766-13
9	18.6	56.4	375	5	PCT-US91-02766-15
10	18.6	56.4	385	6	5428135-3
11	18.6	56.4	649	6	5194396-12
12	18.6	56.4	649	6	5219739-12
13	18.6	56.4	748	1	US-08-387-845-1
14	18.6	56.4	748	2	US-08-788-275-1
15	18.6	56.4	748	3	US-08-867-352-1
16	18.6	56.4	1247	5	PCT-US91-02766-19
17	18.6	56.4	1247	6	5219759-3
18	18.6	56.4	1316	5	PCT-US91-02766-21
19	18	54.5	18609	4	US-08-943-731-1
20	17.6	53.3	924	3	US-08-983-409-3
21	17.4	52.7	600	4	US-09-364-083-1
22	17.4	52.7	639	4	US-09-196-293-10
23	17.4	52.7	639	4	US-08-209-603E-10
24	17.4	52.7	639	4	US-08-233-836C-33
25	17.4	52.7	949	5	PCT-US91-08254-1
26	17.4	52.7	949	5	PCT-US91-08254-2
27	17.4	52.7	1611	1	US-08-061-062A-5

28	17.4	52.7	1611	3	US-08-536-150-5	Sequence 5, Appl
29	17.4	52.7	3250	1	US-08-061-062A-7	Sequence 7, Appl
30	17.4	52.7	3250	3	US-08-536-150-7	Sequence 7, Appl
31	17.4	52.7	5682	2	US-08-663-998-4	Sequence 4, Appl
32	17.2	52.1	31	1	US-08-195-874-3	Sequence 3, Appl
33	17.2	52.1	31	1	US-08-195-874-8	Sequence 8, Appl
34	17.2	52.1	31	5	PCT-US95-01671-3	Sequence 3, Appl
35	17.2	52.1	31	5	PCT-US95-01671-8	Sequence 8, Appl
36	17.2	52.1	2280	2	US-09-055-097-4	Sequence 4, Appl
37	17.2	52.1	7898	1	US-08-984-709A-49	Sequence 49, Appl
38	17	51.5	233	1	US-08-444-142-3	Sequence 3, Appl
39	17	51.5	233	1	US-08-444-131-3	Sequence 3, Appl
40	17	51.5	777	4	US-08-998-416-218	Sequence 218, App
41	17	51.5	1180	2	US-08-867-087B-16	Sequence 16, App
42	17	51.5	1316	6	5219759-1	Patent No. 5219759
43	17	51.5	1987	1	US-07-999-280A-23	Sequence 23, Appl
44	17	51.5	1987	1	US-08-426-036-23	Sequence 23, Appl
45	17	51.5	1987	1	US-08-426-279-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-08-462-484-5
Sequence 5, Application US/08462484
Patent No. 5667531
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Xu, Feng
APPLICANT: Dalborge, Henrik
APPLICANT: Schneider, Palie
APPLICANT: Aaslyng, Dorrit A.
TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 56675310 No. 5667531disk of No. 5667531th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
City: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (ERO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,484
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/441,147
FILING DATE: 15-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lowrey, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4185, 010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 878 9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Polyporus plinthus
FEATURE:
NAME/key: Intron

LOCATION: 666..720
FEATURE:
NAME/KEY: Intron
LOCATION: 790..845
FEATURE:
NAME/KEY: Intron
LOCATION: 1125..1182
FEATURE:
NAME/KEY: Intron
LOCATION: 1390..1450
FEATURE:
NAME/KEY: Intron
LOCATION: 1607..1661
FEATURE:
NAME/KEY: Intron
LOCATION: 1863..1918
FEATURE:
NAME/KEY: Intron
LOCATION: 1976..2025
FEATURE:
NAME/KEY: Intron
LOCATION: 2227..2285
FEATURE:
NAME/KEY: Intron
LOCATION: 2403..2458
FEATURE:
NAME/KEY: Intron
LOCATION: 2576..2627
FEATURE:
NAME/KEY: CDS
LOCATION: 101n (665..721, 789..846, 1124..1183, 1389..1451,
1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
2575..2628).
US-08-462-484-5

Query Match 67.9%; Score 22.4; DB 1; Length 3102;
Best Local Similarity 81.2%; Pred. No. 0.83;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcactctacgcacagaggagactgtaagtc 33
Db 2623 TCATCTTTTGCACAGGAGACTGGAACAAGC 2654

RESULT 2
US-08-441-147-5
Sequence 5, Application US/08441147
Patent No. 5770418
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Xu, Feng
APPLICANT: Dalboge, Henrik
APPLICANT: Schneider, Palle
APPLICANT: Aslyng, Dorrit A.
TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
NUCLEIC ACIDS ENCODING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5770418 No. 5770418 disk of No. 5770418th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,147
FILING DATE: 15-MAY-1995

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4185, 010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 878 9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Polyporus plinthus
FEATURE:
NAME/KEY: Intron
LOCATION: 666..720
FEATURE:
NAME/KEY: Intron
LOCATION: 790..845
FEATURE:
NAME/KEY: Intron
LOCATION: 1125..1182
FEATURE:
NAME/KEY: Intron
LOCATION: 1390..1450
FEATURE:
NAME/KEY: Intron
LOCATION: 1607..1661
FEATURE:
NAME/KEY: Intron
LOCATION: 1863..1918
FEATURE:
NAME/KEY: Intron
LOCATION: 1976..2025
FEATURE:
NAME/KEY: Intron
LOCATION: 2227..2285
FEATURE:
NAME/KEY: Intron
LOCATION: 2403..2458
FEATURE:
NAME/KEY: Intron
LOCATION: 2576..2627
FEATURE:
NAME/KEY: CDS
LOCATION: 101n (665..721, 789..846, 1124..1183, 1389..1451,
1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
2575..2628).
US-08-441-147-5

Query Match 67.9%; Score 22.4; DB 1; Length 3102;
Best Local Similarity 81.2%; Pred. No. 0.83;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcactctacgcacagaggagactgtaagtc 33
Db 2623 TCATCTTTTGCACAGGAGACTGGAACAAGC 2654

RESULT 3
PCT-US95-07536-5
Sequence 5, Application PC/TUS9507536
GENERAL INFORMATION:
APPLICANT:
APPLICANT:
TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
NUCLEIC ACIDS ENCODING SAME
NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Novo Nordisk of North America, Inc.
 STREET: 405 Lexington Avenue, Suite 6400
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/07536
 FILING DATE: 15-June-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/265,534
 FILING DATE: 24-June-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Lowney, Karen A.
 REGISTRATION NUMBER: 31,274
 REFERENCE/DOCKET NUMBER: 4185.204-WO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 867 0123
 TELEFAX: 212 878 9655
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3102 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Polyporus pliusus
 FEATURE:
 NAME/KEY: Intron
 LOCATION: 666..720
 FEATURE:
 NAME/KEY: Intron
 LOCATION: 790..845
 FEATURE:
 NAME/KEY: Intron
 LOCATION: 1125..1182
 FEATURE:
 NAME/KEY: Intron
 LOCATION: 1390..1450
 FEATURE:
 NAME/KEY: Intron
 LOCATION: 1607..1661
 FEATURE:
 NAME/KEY: Intron
 LOCATION: 1863..1918
 FEATURE:
 NAME/KEY: Intron
 LOCATION: 1976..2025
 FEATURE:
 NAME/KEY: Intron
 LOCATION: 2227..2285
 FEATURE:
 NAME/KEY: Intron
 LOCATION: 2403..2458
 FEATURE:
 NAME/KEY: Intron
 LOCATION: 2576..2627
 NAME/KEY: join (665..721, 789..846, 1124..1183, 1389..1451,
 LOCATION: 1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
 LOCATION: 2575..2628).
 PCT-US95-07536-5

Query Match 67.9%; Score 22.4; DB 5; Length 3102;
 Best Local Similarity 81.2%; Pred. No. 0.83;
 Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 2 tcactactgcacagagactgtaagtc 33
 Db 2623 TCATCTTTGACACAGAGACTGGACACAGC 2654
 RESULT 4
 US-08-358-160-143
 Sequence 143, Application US/08358160
 Patent No. 5663143
 GENERAL INFORMATION:
 APPLICANT: LEY, Arthur C.
 APPLICANT: LADNER, Robert C.
 APPLICANT: GUTERMAN, Sonia K.
 APPLICANT: ROBERTS, Bruce L.
 APPLICANT: MARKLAND, William
 APPLICANT: KENT, Rachel B.
 TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
 TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
 NUMBER OF SEQUENCES: 234
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROWDY AND NEIMARK
 STREET: 419 Seventh Street, N.W. Suite 300
 CITY: Washington
 STATE: District of Columbia
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/358,160
 FILING DATE: 16-DEC-1994
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/133,031
 FILING DATE: 13-OCT-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/009,319
 FILING DATE: 26-JAN-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/664,989
 FILING DATE: 01-MAR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/487,063
 FILING DATE: 02-MAR-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/240,160
 FILING DATE: 02-SEP-1988
 ATTORNEY/AGENT INFORMATION:
 NAME: Cooper, Iver P.
 REGISTRATION NUMBER: 28,005
 REFERENCE/DOCKET NUMBER: LEY-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-628-5197
 TELEFAX: 202-737-3528
 TELEX: 248633
 INFORMATION FOR SEQ ID NO: 143:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 234 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: Synthetic DNA fragment
 US-08-358-160-143

Query Match 63.6%; Score 21; DB 1; Length 234;
 Best Local Similarity 82.8%; Pred. No. 1.9;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 tctactgcacagagagactgtaagtc 33
||||| ||||| ||||| ||||| |||||
DB 116 TCTACTACAACTCGAGTATTGGTAAGTGC 144

RESULT 5
US-09-078-294-12/c
Sequence 12, Application US/09078294
Patent No. 6265211
GENERAL INFORMATION:
APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sart, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies Col
CURRENT APPLICATION NUMBER: US/09/078,294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 12
LENGTH: 18073
TYPE: DNA
ORGANISM: BAC-F2 contig 8
US-09-078-294-12

Query Match 61.2% Score 20.2; DB 4; Length 18073;
Best Local Similarity 88.0% Pred. No. 11;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 8 actgcacacagagagactgtaagtc 32
||||| ||||| ||||| ||||| |||||
DB 8535 ACTGTTACAGAGAGACTGTAAATG 8511

RESULT 6
US-08-358-160-145
Sequence 145, Application US/08358160
Patent No. 5663143
GENERAL INFORMATION:
APPLICANT: Ley, Arthur C.
APPLICANT: LADNER, Robert C.
APPLICANT: GUTERMAN, Sonia K.
APPLICANT: ROBERTS, Bruce L.
APPLICANT: MARKLAND, William
APPLICANT: KENT, Rachel B.
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
NUMBER OF SEQUENCES: 234
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W. Suite 300
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,160
FILING DATE: 16-DEC-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,031
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,319
FILING DATE: 26-JAN-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/664,989
FILING DATE: 01-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,063
FILING DATE: 02-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/240,160
FILING DATE: 02-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cooper, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: LEY=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 145:
SEQUENCE CHARACTERISTICS:
LENGTH: 180 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: synthetic DNA fragment
US-08-358-160-145

Query Match 60.0% Score 19.8; DB 1; Length 180;
Best Local Similarity 75.9% Pred. No. 5.8;
Matches 22; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 5 tctactgcacacagagagactgtaagtc 33
:||||| ||||| ||||| ||||| |||||
DB 62 DSTACTACAACTCGAGTATTGGTAAGTGC 90

RESULT 7
US-08-158-353-1/c
Sequence 1, Application US/08158353
Patent No. 5620862
GENERAL INFORMATION:
APPLICANT: Padula, Steven J.
TITLE OF INVENTION: Methods for Diagnosing Early Lyme
TITLE OF INVENTION: Disease
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158,353
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: OCT93-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 824 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

Mon May 13 10:17:29 2002

Search completed: May 12, 2002, 20:14:35
Job time: 7192 sec

us-09-554-024-1.rni

Page 8